A62

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 May 2004 (13.05.2004)

PCT

(10) International Publication Number
WO 2004/039950 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/033610

(22) International Filing Date: 24 October 2003 (24.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/281,341 25 October 2002 (25.10.2002) US

(71) Applicant *(for all designated States except US)*: **CHIRON CORPORATION** [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HOUGHTON, Michael** [GB/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **COATES, Steve** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **SELBY, Mark** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US). **PALIARD, Xavier** [US/US]; c/o Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

(74) Agents: **HARBIN, Alisa** et al.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US).

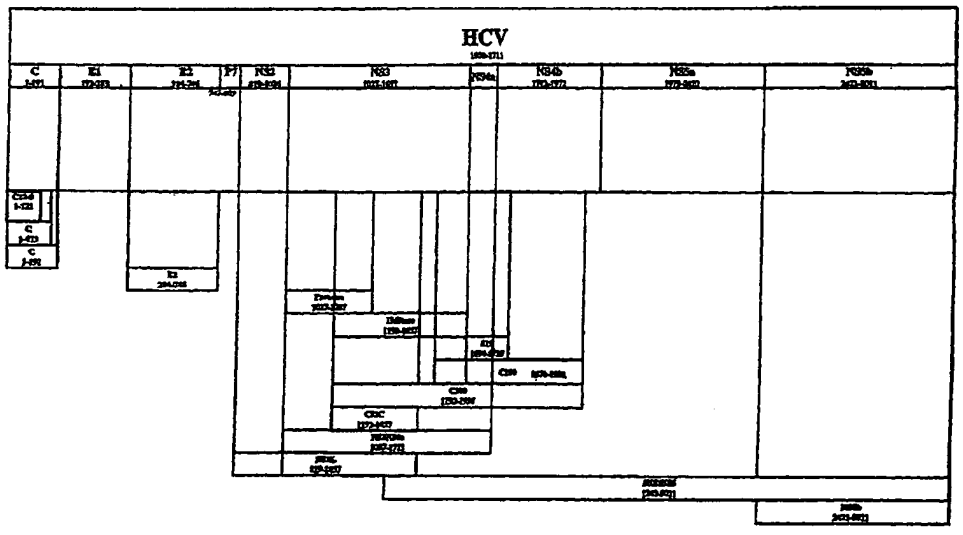
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,

[Continued on next page]

(54) Title: ACTIVATION OF HCV-SPECIFIC CELLS

HCV Genome and Recombinant Proteins



(57) Abstract: The invention provides a method of activating hepatitis C virus (HCV)-specific T cells, including $CD4^+$ and $CD8^+$ T cells. HCV-specific T cells are activated using fusion proteins comprising HCV NS3, NS4, NS5a, and NS5b polypeptides, polynucleotides encoding such fusion proteins, or polypeptide or polynucleotide compositions containing the individual components of these fusions. The method can be used in model systems to develop HCV-specific immunogenic compositions, as well as to immunize a mammal against HCV.

WO 2004/039950 A2

SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ACTIVATION OF HCV-SPECIFIC T CELLS

TECHNICAL AREA OF THE INVENTION

The invention relates to the activation of hepatitis C virus(HCV)-specific T cells. More particularly, the invention relates to the use of multiple HCV polypeptides, either alone or as fusions, to stimulate cell-mediated immune responses, such as to activate HCV-specific T cells.

BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) infection is an important health problem with approximately 1% of the world's population infected with the virus. Over 75% of acutely infected individuals eventually progress to a chronic carrier state that can result in cirrhosis, liver failure, and hepatocellular carcinoma. *See* Alter *et al.* (1992) N. Engl. J. Med. 327:1899-1905; Resnick and Koff. (1993) Arch. Intem. Med. 153:1672-1677; Seeff (1995) Gastrointest. Dis. 6:20-27; Tong *et al.* (1995) N. Engl. J. Med. 332:1463-1466.

Despite extensive advances in the development of pharmaceuticals against certain viruses like HIV, control of acute and chronic HCV infection has had limited success (Hoofnagle and di Bisceglie (1997) N. Engl. J. Med. 336:347-356). In particular, generation of a strong cytotoxic T lymphocyte (CTL) response is thought to be important for the control and eradication of HCV infections. Thus, there is a need in the art for effective methods of inducing strong CTL responses against HCV.

SUMMARY OF THE INVENTION

It is an object of the invention to provide reagents and methods for stimulating immune responses, such as activating T cells which recognize epitopes of HCV polypeptides. This and other objects of the invention are provided by one or more of the embodiments described below.

The invention provides HCV proteins useful for stimulating immune responses, such as activating HCV-specific T cells. One embodiment provides a fusion protein that comprises HCV polypeptides, wherein the HCV polypeptides consist essentially of an NS3, an NS4, an NS5a polypeptide, and optionally a core polypeptide. In certain embodiments, the fusion protein includes an NS5b polypeptide.

In certain embodiments, at least one of the HCV polypeptides is derived from a different strain of HCV than the other polypeptides.

The invention also provides compositions comprising any of these fusion proteins and a pharmaceutically acceptable excipient. In certain embodiments, the compositions further comprise an adjuvant, a CpG polynucleotide and/or the fusion protein is adsorbed to or entrapped within a microparticle or ISCOM. The compositions can further comprise a polynucleotide encoding an E1E2 complex. The E1E2 polynucleotide can also be adsorbed to or entrapped withing a microparticle.

Another embodiment provides a composition comprising HCV polypeptides and a pharmaceutically acceptable excipient. The HCV polypeptides consist essentially of an NS3, an NS4, an NS5a polypeptide, and optionally a core polypeptide. In certain embodiments, the composition includes an NS5b polypeptide. In other embodiments, the compositions further comprise an adjuvant, a CpG polynucleotide and/or one or more of the HCV polypeptides is adsorbed to or entrapped within a microparticle or ISCOM. The compositions can further comprise a polynucleotide encoding an E1E2 complex. The E1E2 polynucleotide can also be adsorbed to or entrapped withing a microparticle.

Moreover, one of the HCV polypeptides may be derived from a different strain of HCV than the others.

Even another embodiment of the invention provides an isolated and purified polynucleotide which encodes a fusion protein as described above. In additional embodiments, the fusion proteins further include a polynucleotide encoding an E1E2 complex.

Yet another embodiment of the invention provides a composition comprising the polynucleotides described above and a pharmaceutically acceptable excipient. In certain embodiments, the compositions further comprise an adjuvant and/or the polynucleotide may be adsorbed to or entrapped within a microparticle. The compositions can further comprise a polynucleotide encoding an E1E2 complex. The E1E2 polynucleotide can also be adsorbed to or entrapped withing a microparticle.

In a further embodiment, the invention provides a composition comprising HCV polynucleotides and a pharmaceutically acceptable excipient, wherein the HCV polynucleotides consist essentially of polynucleotides encoding an NS3, an NS4, an NS5a polypeptide, and optionally a core polypeptide. In certain embodiments, the composition also includes a polynucleotide encoding an NS5b polypeptide. The compositions may further comprise an adjuvant and/or one or more of the polynucleotides may be adsorbed to or entrapped within a microparticle. The compositions can further comprise a polynucleotide encoding an E1E2 complex. The E1E2 polynucleotide can also be adsorbed to or entrapped withing a microparticle. Additionally, one or more of the polynucleotides may be derived from a different strain of HCV than the others.

In another embodiment, the invention provides a method of activating T cells which recognize an epitope of an HCV polypeptide. T cells are contacted with any of the fusions, polynucleotides or compositions described above. A population of activated T cells recognizes an epitope of the NS3, NS4, NS5a, NS5b, core and/or E1E2 polypeptide.

In the proteins and polynucleotides above, the regions in the fusions need not be in the order in which they naturally occur in the native HCV polyprotein. Thus, for example, the NS5b polypeptide, if present, may be at the N– and/or C-terminus of the fusion, or may be located internally. Similarly, the E1 polypeptide may precede or follow the E2 polypeptide. The E1E2 polypeptide may also be part of the

nonstructural fusion protein or may be provided separately, as an E1E2 complex, or as individual polypeptides.

Moreover, the NS3 polypeptide may include a modification to inhibit protease activity, such that cleavage of the fusion is inhibited. Such modifications are described more fully below. Additionally, the compositions can comprise more than one HCV nonstructural fusion protein, such as a fusion protein with NS3, NS4 and NS5a, and a fusion protein with NS3, NS4, NS5a, NS5b and E1E2. The E1E2 complexes, whether present separately or as part of the fusion, can have varying E1E2 polypeptides (described more fully below).

In certain embodiments, the nonstructural fusion protein consists of, from the amino terminus to the carboxyl terminus, an NS3, an NS4, an NS5a and, optionally, an NS5b polypeptide and the E1E2 complex consists of, from amino terminus to the carboxyl terminus, an E1 polypeptide and an E2 polypeptide.

The various polypeptides (and polynucleotides encoding therefor) are derived from the same HCV isolate, or from different strains and isolates including isolates having any of the various HCV genotypes, to provide increased protection against a broad range of HCV genotypes.

Yet another embodiment of the invention provides a method of stimulating an immune response, such as a cellular immune response, in a vertebrate subject by administering a composition as described herein. In certain embodiments, the composition activates T cells which recognize an epitope of an HCV polypeptide. T cells are contacted with a composition as described above. A population of activated T cells recognizes an epitope of one or more of the HCV polypeptide(s).

The invention thus provides methods and reagents for stimulating immune responses to HCV, such as for activating T cells which recognize epitopes of HCV polypeptides. These methods and reagents are particularly advantageous for identifying epitopes of HCV polypeptides associated with a strong CTL response and for immunizing mammals, including humans, against HCV.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagrammatic representation of the HCV genome, depicting the various regions of the HCV polyprotein.

Figure 2 (SEQ ID NOS: 9 and 10) depicts the DNA and corresponding amino acid sequence of a representative native NS3 protease domain.

Figures 3A-3C (SEQ ID NOS:3 and 4) shows the nucleotide and corresponding amino acid sequence for the HCV-1 E1/E2/p7 region. The numbers shown in the figure are relative to the full-length HCV-1 polyprotein. The E1, E2 and p7 regions are shown.

Figure 4 is a diagram of plasmid pMHE1E2-809, encoding $E1E2_{809}$, a representative E1E2 protein for use with the present invention.

Figures 5A-5J (SEQ ID NOS:7 and 8) depict the DNA and corresponding amino acid sequence of a representative NS345Core fusion protein. The depicted sequence includes amino acids 1242-3011 of the HCV polyprotein (representing polypeptides from NS3, NS4, NS5a and NS5b) with amino acids 1-121 of the HCV polyprotein (representing a polypeptide from the core region) fused to the C-terminus of NS5b. This numbering is relative to the HCV-1 polyprotein.

Figure 6 shows a side-by-side comparison of IFN-γ expression generated in animals in response to delivery of alphavirus constructs encoding NS3NS4NS5a.

Figure 7 shows IFN-γ expression generated in animals in response to delivery of plasmid DNA encoding NS3NS4NS5a ("naked"), PLG-linked DNA encoding NS3NS4NS5a ("PLG), separate DNA plasmids encoding NS5a, NS34a, and NS4ab ("naked"), and PLG-linked DNA encoding NS5a, NS34a, and NS4ab ("PLG").

Figure 8 shows HCV-specific CD8+ and CD4+ responses in vaccinated chimpanzees.

Figure 9 depicts the specificity of T cell responses primed by electroporation of plasmid DNA two weeks subsequent to the third immunization.

Figure 10 shows the specificity of T cell responses primed by vaccinating chimpanzees with $NS345Core_{121}$-ISCOMS two weeks subsequent to the third immunization.

DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, recombinant DNA techniques and immunology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *DNA Cloning*, Vols. I and II (D.N. Glover ed.); *Oligonucleotide Synthesis* (M.J. Gait ed.); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds.); *Animal Cell Culture* (R.K. Freshney ed.); Perbal, B., *A Practical Guide to Molecular Cloning.*

It must be noted that, as used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more antigens, and the like.

The following amino acid abbreviations are used throughout the text:

Alanine: Ala (A)	Arginine: Arg (R)
Asparagine: Asn (N)	Aspartic acid: Asp (D)
Cysteine: Cys (C)	Glutamine: Gln (Q)
Glutamic acid: Glu (E)	Glycine: Gly (G)
Histidine: His (H)	Isoleucine: Ile (I)
Leucine: Leu (L)	Lysine: Lys (K)
Methionine: Met (M)	Phenylalanine: Phe (F)
Proline: Pro (P)	Serine: Ser (S)
Threonine: Thr (T)	Tryptophan: Trp (W)
Tyrosine: Tyr (Y)	Valine: Val (V)

I. Definitions

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a minimum length of the product. Thus, peptides, oligopeptides, dimers, multimers, and the like, are included within the definition.

Both full-length proteins and fragments thereof are encompassed by the definition. The terms also include postexpression modifications of the polypeptide, for example, glycosylation, acetylation, phosphorylation and the like. Furthermore, for purposes of the present invention, a "polypeptide" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the desired activity. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the proteins or errors due to PCR amplification.

An HCV polypeptide is a polypeptide, as defined above, derived from the HCV polyprotein. The polypeptide need not be physically derived from HCV, but may be synthetically or recombinantly produced. Moreover, the polypeptide may be derived from any of the various HCV strains and isolates including isolates having any of the 6 genotypes of HCV described in Simmonds et al., *J. Gen. Virol.* (1993) 74:2391-2399 (e.g., strains 1, 2, 3, 4 etc.), as well as newly identified isolates, and subtypes of these isolates, such as HCV1a, HCV1b, etc. A number of conserved and variable regions are known between these strains and, in general, the amino acid sequences of epitopes derived from these regions will have a high degree of sequence homology, e.g., amino acid sequence homology of more than 30%, preferably more than 40%, when the two sequences are aligned. Thus, for example, the term "NS4" polypeptide refers to native NS4 from any of the various HCV strains, as well as NS4 analogs, muteins and immunogenic fragments, as defined further below.

By an "E1 polypeptide" is meant a molecule derived from an HCV E1 region. The mature E1 region of HCV-1 begins at approximately amino acid 192 of the polyprotein and continues to approximately amino acid 383, numbered relative to the full-length HCV-1 polyprotein. (See, Figures 1 and 3A-3C. Amino acids 192-383 of Figures 3A-3C correspond to amino acid positions 20-211 of SEQ ID NO:4.) Amino acids at around 173 through approximately 191 (amino acids 1-19 of SEQ ID NO: 4) serve as a signal sequence for E1. Thus, by an "E1 polypeptide" is meant either a precursor E1 protein, including the signal sequence, or a mature E1 polypeptide which lacks this sequence, or even an E1 polypeptide with a heterologous signal sequence.

The E1 polypeptide includes a C-terminal membrane anchor sequence which occurs at approximately amino acid positions 360-383 (see, International Publication No. WO 96/04301, published February 15, 1996). An E1 polypeptide, as defined herein, may or may not include the C-terminal anchor sequence or portions thereof.

By an "E2 polypeptide" is meant a molecule derived from an HCV E2 region. The mature E2 region of HCV-1 begins at approximately amino acid 383-385, numbered relative to the full-length HCV-1 polyprotein. (See, Figures 1 and 3A-3C. Amino acids 383-385 of Figures 3A-3C correspond to amino acid positions 211-213 of SEQ ID NO:4.) A signal peptide begins at approximately amino acid 364 of the polyprotein. Thus, by an "E2 polypeptide" is meant either a precursor E2 protein, including the signal sequence, or a mature E2 polypeptide which lacks this sequence, or even an E2 polypeptide with a heterologous signal sequence. The E2 polypeptide includes a C-terminal membrane anchor sequence which occurs at approximately amino acid positions 715-730 and may extend as far as approximately amino acid residue 746 (see, Lin et al., *J. Virol.* (1994) 68:5063-5073). An E2 polypeptide, as defined herein, may or may not include the C-terminal anchor sequence or portions thereof. Moreover, an E2 polypeptide may also include all or a portion of the p7 region which occurs immediately adjacent to the C-terminus of E2. As shown in Figures 1 and 3A-3C, the p7 region is found at positions 747-809, numbered relative to the full-length HCV-1 polyprotein (amino acid positions 575-637 of SEQ ID NO:4). Additionally, it is known that multiple species of HCV E2 exist (Spaete et al., *Virol.* (1992) 188:819-830; Selby et al., *J. Virol.* (1996) 70:5177-5182; Grakoui et al., *J. Virol.* (1993) 67:1385-1395; Tomei et al., *J. Virol.* (1993) 67:4017-4026). Accordingly, for purposes of the present invention, the term "E2" encompasses any of these species of E2 including, without limitation, species that have deletions of 1-20 or more of the amino acids from the N-terminus of the E2, such as, e.g, deletions of 1, 2, 3, 4, 5....10...15, 16, 17, 18, 19... etc. amino acids. Such E2 species include those beginning at amino acid 387, amino acid 402, amino acid 403, etc.

Representative E1 and E2 regions from HCV-1 are shown in Figures 3A-3C and SEQ ID NO:4. For purposes of the present invention, the E1 and E2 regions are defined with respect to the amino acid number of the polyprotein encoded by the

genome of HCV-1, with the initiator methionine being designated position 1. See, e.g., Choo et al., *Proc. Natl. Acad. Sci. USA* (1991) 88:2451-2455. However, it should be noted that the term an "E1 polypeptide" or an "E2 polypeptide" as used herein is not limited to the HCV-1 sequence. In this regard, the corresponding E1 or E2 regions in other HCV isolates can be readily determined by aligning sequences from the isolates in a manner that brings the sequences into maximum alignment. This can be performed with any of a number of computer software packages, such as ALIGN 1.0, available from the University of Virginia, Department of Biochemistry (Attn: Dr. William R. Pearson). See, Pearson et al., *Proc. Natl. Acad. Sci. USA* (1988) 85:2444-2448.

Furthermore, an "E1 polypeptide" or an "E2 polypeptide" as defined herein is not limited to a polypeptide having the exact sequence depicted in the Figures. Indeed, the HCV genome is in a state of constant flux *in vivo* and contains several variable domains which exhibit relatively high degrees of variability between isolates. A number of conserved and variable regions are known between these strains and, in general, the amino acid sequences of epitopes derived from these regions will have a high degree of sequence homology, e.g., amino acid sequence homology of more than 30%, preferably more than 40%, more than 60%, and even more than 80-90% homology, when the two sequences are aligned. It is readily apparent that the terms encompass E1 and E2 polypeptides from any of the various HCV strains and isolates including isolates having any of the 6 genotypes of HCV described in Simmonds et al., *J. Gen. Virol.* (1993) 74:2391-2399 (e.g., strains 1, 2, 3, 4 etc.), as well as newly identified isolates, and subtypes of these isolates, such as HCV1a, HCV1b etc.

Thus, for example, the term "E1" or "E2" polypeptide refers to native E1 or E2 sequences from any of the various HCV strains, as well as analogs, muteins and immunogenic fragments, as defined further below. The complete genotypes of many of these strains are known. See, e.g., U.S. Patent No. 6,150,087 and GenBank Accession Nos. AJ238800 and AJ238799.

Additionally, the terms "E1 polypeptide" and "E2 polypeptide" encompass proteins which include modifications to the native sequence, such as internal deletions, additions and substitutions (generally conservative in nature). These

modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through naturally occurring mutational events. All of these modifications are encompassed in the present invention so long as the modified E1 and E2 polypeptides function for their intended purpose. Thus, for example, if the E1 and/or E2 polypeptides are to be used in vaccine compositions, the modifications must be such that immunological activity (i.e., the ability to elicit a humoral or cellular immune response to the polypeptide) is not lost.

By "E1E2" complex is meant a protein containing at least one E1 polypeptide and at least one E2 polypeptide, as described above. Such a complex may also include all or a portion of the p7 region which occurs immediately adjacent to the C-terminus of E2. As shown in Figures 1 and 3A-3C, the p7 region is found at positions 747-809, numbered relative to the full-length HCV-1 polyprotein (amino acid positions 575-637 of SEQ ID NO:4). A representative E1E2 complex which includes the p7 protein is termed "$E1E2_{809}$" herein.

The mode of association of E1 and E2 in an E1E2 complex is immaterial. The E1 and E2 polypeptides may be associated through non-covalent interactions such as through electrostatic forces, or by covalent bonds. For example, the E1E2 polypeptides of the present invention may be in the form of a fusion protein which includes an immunogenic E1 polypeptide and an immunogenic E2 polypeptide, as defined above. The fusion may be expressed from a polynucleotide encoding an E1E2 chimera. Alternatively, E1E2 complexes may form spontaneously simply by mixing E1 and E2 proteins which have been produced individually. Similarly, when co-expressed and secreted into media, the E1 and E2 proteins can form a complex spontaneously. Thus, the term encompasses E1E2 complexes (also called aggregates) that spontaneously form upon purification of E1 and/or E2. Such aggregates may include one or more E1 monomers in association with one or more E2 monomers. The number of E1 and E2 monomers present need not be equal so long as at least one E1 monomer and one E2 monomer are present. Detection of the presence of an E1E2 complex is readily determined using standard protein detection techniques such as polyacrylamide gel electrophoresis and immunological techniques such as immunoprecipitation.

The terms "analog" and "mutein" refer to biologically active derivatives of the reference molecule, or fragments of such derivatives, that retain desired activity, such as the ability to stimulate a cell-mediated immune response, as defined below. In general, the term "analog" refers to compounds having a native polypeptide sequence and structure with one or more amino acid additions, substitutions (generally conservative in nature) and/or deletions, relative to the native molecule, so long as the modifications do not destroy immunogenic activity. The term "mutein" refers to peptides having one or more peptide mimics ("peptoids"), such as those described in International Publication No. WO 91/04282. Preferably, the analog or mutein has at least the same immunoactivity as the native molecule. Methods for making polypeptide analogs and muteins are known in the art and are described further below.

As explained above, analogs generally include substitutions that are conservative in nature, i.e., those substitutions that take place within a family of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cysteine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. For example, it is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the biological activity. For example, the polypeptide of interest may include up to about 5-10 conservative or non-conservative amino acid substitutions, or even up to about 15-25 conservative or non-conservative amino acid substitutions, or any integer between 5-25, so long as the desired function of the molecule remains intact. One of skill in the art may readily determine regions of the molecule of interest that can tolerate change by reference to Hopp/Woods and Kyte-Doolittle plots, well known in the art.

By "modified NS3" is meant an NS3 polypeptide with a modification such that protease activity of the NS3 polypeptide is disrupted. The modification can include one or more amino acid additions, substitutions (generally non-conservative in nature) and/or deletions, relative to the native molecule, wherein the protease activity of the NS3 polypeptide is disrupted. Methods of measuring protease activity are discussed further below.

By "fragment" is intended a polypeptide consisting of only a part of the intact full-length polypeptide sequence and structure. The fragment can include a C-terminal deletion and/or an N-terminal deletion of the native polypeptide. An "immunogenic fragment" of a particular HCV protein will generally include at least about 5-10 contiguous amino acid residues of the full-length molecule, preferably at least about 15-25 contiguous amino acid residues of the full-length molecule, and most preferably at least about 20-50 or more contiguous amino acid residues of the full-length molecule, that define an epitope, or any integer between 5 amino acids and the full-length sequence, provided that the fragment in question retains immunogenic activity, as measured by the assays described herein.

The term "epitope" as used herein refers to a sequence of at least about 3 to 5, preferably about 5 to 10 or 15, and not more than about 1,000 amino acids (or any integer therebetween), which define a sequence that by itself or as part of a larger sequence, binds to an antibody generated in response to such sequence. There is no critical upper limit to the length of the fragment, which may comprise nearly the full-length of the protein sequence, or even a fusion protein comprising two or more epitopes from the HCV polyprotein. An epitope for use in the subject invention is not limited to a polypeptide having the exact sequence of the portion of the parent protein from which it is derived. Indeed, viral genomes are in a state of constant flux and contain several variable domains which exhibit relatively high degrees of variability between isolates. Thus the term "epitope" encompasses sequences identical to the native sequence, as well as modifications to the native sequence, such as deletions, additions and substitutions (generally conservative in nature).

Regions of a given polypeptide that include an epitope can be identified using any number of epitope mapping techniques, well known in the art. See, e.g., *Epitope*

Mapping Protocols in Methods in Molecular Biology, Vol. 66 (Glenn E. Morris, Ed., 1996) Humana Press, Totowa, New Jersey. For example, linear epitopes may be determined by e.g., concurrently synthesizing large numbers of peptides on solid supports, the peptides corresponding to portions of the protein molecule, and reacting the peptides with antibodies while the peptides are still attached to the supports. Such techniques are known in the art and described in, e.g., U.S. Patent No. 4,708,871; Geysen et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:3998-4002; Geysen et al. (1986) *Molec. Immunol.* 23:709-715. Similarly, conformational epitopes are readily identified by determining spatial conformation of amino acids such as by, e.g., x-ray crystallography and 2-dimensional nuclear magnetic resonance. See, e.g., *Epitope Mapping Protocols, supra.* Antigenic regions of proteins can also be identified using standard antigenicity and hydropathy plots, such as those calculated using, e.g., the Omiga version 1.0 software program available from the Oxford Molecular Group. This computer program employs the Hopp/Woods method, Hopp et al., *Proc. Natl. Acad. Sci USA* (1981) 78:3824-3828 for determining antigenicity profiles, and the Kyte-Doolittle technique, Kyte et al., *J. Mol. Biol.* (1982) 157:105-132 for hydropathy plots.

For a description of various HCV epitopes, see, e.g., Chien et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:10011-10015; Chien et al., *J. Gastroent. Hepatol.* (1993) 8:S33-39; Chien et al., International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; and U.S. Patent Nos. 6,280,927 and 6,150,087.

As used herein, the term "conformational epitope" refers to a portion of a full-length protein, or an analog or mutein thereof, having structural features native to the amino acid sequence encoding the epitope within the full-length natural protein. Native structural features include, but are not limited to, glycosylation and three dimensional structure. Preferably, a conformational epitope is produced recombinantly and is expressed in a cell from which it is extractable under conditions which preserve its desired structural features, e.g. without denaturation of the epitope. Such cells include bacteria, yeast, insect, and mammalian cells. Expression and isolation of recombinant conformational epitopes from the HCV polyprotein are

described in e.g., International Publication Nos. WO 96/04301, WO 94/01778, WO 95/33053, WO 92/08734.

As used herein the term "T-cell epitope" refers to a feature of a peptide structure which is capable of inducing T-cell immunity towards the peptide structure or an associated hapten. T-cell epitopes generally comprise linear peptide determinants that assume extended conformations within the peptide-binding cleft of MHC molecules, (Unanue et al., *Science* (1987) 236:551-557). Conversion of polypeptides to MHC class II-associated linear peptide determinants (generally between 5-14 amino acids in length) is termed "antigen processing" which is carried out by antigen presenting cells (APCs). More particularly, a T-cell epitope is defined by local features of a short peptide structure, such as primary amino acid sequence properties involving charge and hydrophobicity, and certain types of secondary structure, such as helicity, that do not depend on the folding of the entire polypeptide. Further, it is believed that short peptides capable of recognition by helper T-cells are generally amphipathic structures comprising a hydrophobic side (for interaction with the MHC molecule) and a hydrophilic side (for interacting with the T-cell receptor), (Margalit et al., *Computer Prediction of T-cell Epitopes*, New Generation Vaccines Marcel-Dekker, Inc, ed. G.C. Woodrow et al., (1990) pp. 109-116) and further that the amphipathic structures have an α-helical configuration (see, e.g., Spouge et al., *J. Immunol.* (1987) 138:204-212; Berkower et al., *J. Immunol.* (1986) 136:2498-2503).

Hence, segments of proteins that include T-cell epitopes can be readily predicted using numerous computer programs. (See e.g., Margalit et al., *Computer Prediction of T-cell Epitopes*, New Generation Vaccines Marcel-Dekker, Inc, ed. G.C. Woodrow et al., (1990) pp. 109-116). Such programs generally compare the amino acid sequence of a peptide to sequences known to induce a T-cell response, and search for patterns of amino acids which are believed to be required for a T-cell epitope.

An "immunological response" to an HCV antigen (including both polypeptide and polynucleotides encoding polypeptides that are expressed *in vivo*) or composition is the development in a subject of a humoral and/or a cellular immune response to molecules present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to an immune response mediated by

antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTLs"). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the intracellular destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J. Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J. Immunol.* (1994) 24:2369-2376; and the examples below.

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the activation of suppressor T-cells and/or

γδ T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell cytotoxicity (ADCC) to provide protection or alleviation of symptoms to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

By "equivalent antigenic determinant" is meant an antigenic determinant from different sub-species or strains of HCV, such as from strains 1, 2, 3, etc., of HCV which antigenic determinants are not necessarily identical due to sequence variation, but which occur in equivalent positions in the HCV sequence in question. In general the amino acid sequences of equivalent antigenic determinants will have a high degree of sequence homology, e.g., amino acid sequence homology of more than 30%, usually more than 40%, such as more than 60%, and even more than 80-90% homology, when the two sequences are aligned.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A transcription termination sequence may be located 3' to the coding sequence.

A "nucleic acid" molecule or "polynucleotide" can include both double- and single-stranded sequences and refers to, but is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral (e.g. DNA viruses and retroviruses) or procaryotic DNA, and especially synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

An "HCV polynucleotide" is a polynucleotide that encodes an HCV polypeptide, as defined above.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their desired function. Thus,

a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper transcription factors, etc., are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence, as can transcribed introns, and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, viral, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation is not associated with all or a portion of the polynucleotide with which it is associated in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. In general, the gene of interest is cloned and then expressed in transformed organisms, as described further below. The host organism expresses the foreign gene to produce the protein under expression conditions.

A "control element" refers to a polynucleotide sequence which aids in the expression of a coding sequence to which it is linked. The term includes promoters, transcription termination sequences, upstream regulatory domains, polyadenylation signals, untranslated regions, including 5'-UTRs and 3'-UTRs and when appropriate, leader sequences and enhancers, which collectively provide for the transcription and translation of a coding sequence in a host cell.

A "promoter" as used herein is a DNA regulatory region capable of binding RNA polymerase in a host cell and initiating transcription of a downstream (3' direction) coding sequence operably linked thereto. For purposes of the present invention, a promoter sequence includes the minimum number of bases or elements necessary to initiate transcription of a gene of interest at levels detectable above background. Within the promoter sequence is a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eucaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

A control sequence "directs the transcription" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

"Expression cassette" or "expression construct" refers to an assembly which is capable of directing the expression of the sequence(s) or gene(s) of interest. The expression cassette includes control elements, as described above, such as a promoter which is operably linked to (so as to direct transcription of) the sequence(s) or gene(s) of interest, and often includes a polyadenylation sequence as well. Within certain embodiments of the invention, the expression cassette described herein may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include, one or more selectable markers, a signal which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), at least one multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

"Transformation," as used herein, refers to the insertion of an exogenous polynucleotide into a host cell, irrespective of the method used for insertion: for example, transformation by direct uptake, transfection, infection, and the like. For particular methods of transfection, see further below. The exogenous polynucleotide may be maintained as a nonintegrated vector, for example, an episome, or alternatively, may be integrated into the host genome.

A "host cell" is a cell which has been transformed, or is capable of transformation, by an exogenous DNA sequence.

By "isolated" is meant, when referring to a polypeptide, that the indicated molecule is separate and discrete from the whole organism with which the molecule is found in nature or is present in the substantial absence of other biological macro-molecules of the same type. The term "isolated" with respect to a polynucleotide is a nucleic acid molecule devoid, in whole or part, of sequences normally associated with it in nature; or a sequence, as it exists in nature, but having heterologous sequences in association therewith; or a molecule disassociated from the chromosome.

The term "purified" as used herein preferably means at least 75% by weight, more preferably at least 85% by weight, more preferably still at least 95% by weight, and most preferably at least 98% by weight, of biological macromolecules of the same type are present.

"Homology" refers to the percent identity between two polynucleotide or two polypeptide moieties. Two DNA, or two polypeptide sequences are "substantially homologous" to each other when the sequences exhibit at least about 50% , preferably at least about 75%, more preferably at least about 80%-85%, preferably at least about 90%, and most preferably at least about 95%-98%, or more, sequence identity over a defined length of the molecules. As used herein, substantially homologous also refers to sequences showing complete identity to the specified DNA or polypeptide sequence.

In general, "identity" refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Percent identity can be determined by a direct comparison of the sequence information between two molecules by aligning the sequences, counting the exact number of matches between the two aligned sequences, dividing by the length of the shorter sequence, and multiplying the result by 100. Readily available computer programs can be used to aid in the analysis, such as ALIGN, Dayhoff, M.O. in *Atlas of Protein Sequence and Structure* M.O. Dayhoff ed., 5 Suppl. 3:353-358, National biomedical Research Foundation, Washington, DC, which adapts the local homology algorithm of Smith and Waterman *Advances in Appl. Math.* 2:482-489, 1981 for peptide analysis. Programs for determining nucleotide sequence identity are available in the Wisconsin Sequence Analysis Package, Version 8 (available from Genetics Computer Group, Madison, WI) for example, the BESTFIT, FASTA and GAP programs, which also rely on the Smith and Waterman algorithm. These programs are readily utilized with the default parameters recommended by the manufacturer and described in the Wisconsin Sequence Analysis Package referred to above. For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and Waterman with a default scoring table and a gap penalty of six nucleotide positions.

Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by = HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at the following internet address: http://www.ncbi.nlm.gov/cgi-bin/BLAST.

Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s), and size determination of the digested fragments. DNA sequences that are substantially homologous can be identified in a Southern hybridization experiment under, for example, stringent conditions, as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al., *supra*; *DNA Cloning, supra*; *Nucleic Acid Hybridization, supra.*

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of the antigen or antigens. The nucleic acid molecule can be introduced directly into the recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced

into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may be effected prophylactically (prior to infection) or therapeutically (following infection).

By "vertebrate subject" is meant any member of the subphylum cordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The invention described herein is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

II. Modes of Carrying out the Invention

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of compositions and methods similar or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

It is a discovery of the present invention that fusion proteins, combinations of the individual components of these fusions, and polynucleotides encoding the same, comprising an NS3, an NS4, and an NS5a polypeptide with or without a core polypeptide, or an NS3, an NS4, an NS5a, and an NS5b polypeptide, with or without a

core polypeptide, of an HCV virus can be used to activate HCV-specific T cells, i.e., T cells which recognize epitopes of these polypeptides.

The present invention also pertains to compositions comprising HCV nonstructural fusion proteins and HCV E1E2 complexes, as well as compositions comprising polynucleotides encoding the same or combinations of polypeptides and polynucleotides.

The proteins, polynucleotides, compositions and combinations of the present invention can be used to stimulate a cellular immune response, such as to activate HCV-specific T cells, i.e., T cells which recognize epitopes of these polypeptides. Activation of HCV-specific T cells provides both *in vitro* and *in vivo* model systems for the development of HCV vaccines, particularly for identifying HCV polypeptide epitopes associated with a response. The compositions can also be used to generate an immune response against HCV in a mammal, particularly a CTL response for either therapeutic or prophylactic purposes.

Fusion Proteins

The genomes of HCV strains contain a single open reading frame of approximately 9,000 to 12,000 nucleotides, which is transcribed into a polyprotein. As shown in Figure 1 and the table below, an HCV polyprotein, upon cleavage, produces at least ten distinct products, in the order of NH_2- Core-E1-E2-p7-NS2-NS3-NS4a-NS4b-NS5a-NS5b-COOH. The core polypeptide occurs at positions 1-191, numbered relative to HCV-1 (see, Choo et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:2451-2455, for the HCV-1 genome). This polypeptide is further processed to produce an HCV polypeptide with approximately amino acids 1-173. The envelope polypeptides, E1 and E2, occur at about positions 192-383 and 384-746, respectively. The P7 domain is found at about positions 747-809. NS2 is an integral membrane protein with proteolytic activity and is found at about positions 810-1026 of the polyprotein. NS2, in combination with NS3, (found at about positions 1027-1657), cleaves the NS2-NS3 sissle bond which in turn generates the NS3 N-terminus and releases a large polyprotein that includes both serine protease and RNA helicase activities. The NS3 protease, found at about positions 1027-1207, serves to process

the remaining polyprotein. The helicase activity is found at about positions 1193-1657. NS3 liberates an NS3 cofactor (NS4a, found about positions 1658-1711), two proteins (NS4b found at about positions 1712-1972, and NS5a found at about positions 1973-2420), and an RNA-dependent RNA polymerase (NS5b found at about positions 2421-3011). Completion of polyprotein maturation is initiated by autocatalytic cleavage at the NS3-Ns4a junction, catalyzed by the NS3 serine protease.

Domain	Approximate Boundaries*
C (core)	1-191
E1	192-383
E2	384-746
P7	747-809
NS2	810-1026
NS3	1027-1657
NS4a	1658-1711
NS4b	1712-1972
NS5a	1973-2420
NS5b	2421-3011

*Numbered relative to HCV-1. See, Choo et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:2451-2455.

Fusion proteins for use in the compositions and methods, and polynucleotides encoding therefor, include or encode an NS3 polypeptide, an NS4 (NS4a and/or NS4b) polypeptide, an NS5a polypeptide and, optionally, an NS5b polypeptide. The fusion proteins may or may not include all or part of the core region. In certain embodiments, none of the core region is present in the compositions. The nonstructural regions need not be in the order in which they naturally occur in the native HCV polyprotein. Thus, for example, the NS5b polypeptide may be at the N- and/or C-terminus of the fusion or may be found internally. These polypeptides may

be derived from the same HCV isolate, or from different strains and isolates including isolates having any of the various HCV genotypes, to provide increased protection against a broad range of HCV genotypes. Additionally, polypeptides can be selected based on the particular viral clades endemic in specific geographic regions where vaccine compositions containing the fusions will be used. It is readily apparent that the subject fusions provide an effective means of treating HCV infection in a wide variety of contexts.

In one embodiment, the fusion protein of the present invention includes an NS3 polypeptide that has been modified to inhibit protease activity, such that further cleavage of the fusion is inhibited. The NS3 polypeptide can be modified by deletion of all or a portion of the NS3 protease domain. Alternatively, proteolytic activity can be inhibited by substitutions of amino acids within active regions of the protease domain. Finally, additions of amino acids to active regions of the domain, such that the catalytic site is modified, will also serve to inhibit proteolytic activity.

As explained above, the protease activity is found at about amino acid positions 1027-1207, numbered relative to the full-length HCV-1 polyprotein (see, Choo et al., *Proc. Natl. Acad. Sci. USA* (1991) 88:2451-2455), positions 2-182 of Figure 3. The structure of the NS3 protease and active site are known. See, e.g., De Francesco et al., *Antivir. Ther.* (1998) 3:99-109; Koch et al., *Biochemistry* (2001) 40:631-640. Thus, deletions or modifications to the native sequence will typically occur at or near the active site of the molecule. Particularly, it is desirable to modify or make deletions to one or more amino acids occurring at positions 1- or 2-182, preferably 1- or 2-170, or 1- or 2-155 of Figure 3. Preferred modifications are to the catalytic triad at the active site of the protease, i.e., H, D or S residues, in order to inactivate the protease. These residues occur at positions 1083, 1105 and 1165, respectively, numbered relative to the full-length HCV polyprotein (positions 58, 80 and 140, respectively, of Figure 3). Such modifications will suppress proteolytic cleavage while maintaining T-cell epitopes.

One of skill in the art can readily determine portions of the NS3 protease to delete in order to disrupt activity. The presence or absence of activity can be determined using methods known to those of skill in the art.

For example, protease activity or lack thereof may be determined using assays well known in the art. See, e.g., Takeshita et al., *Anal. Biochem.* (1997) 247:242-246; Kakiuchi et al., *J. Biochem.* (1997) 122:749-755; Sali et al., *Biochemistry* (1998) 37:3392-3401; Cho et al., *J. Virol. Meth.* (1998) 72:109-115; Cerretani et al., *Anal. Biochem.* (1999) 266:192-197; Zhang et al., *Anal. Biochem.* (1999) 270:268-275; Kakiuchi et al., *J. Virol. Meth.* (1999) 80:77-84; Fowler et al., *J. Biomol. Screen.* (2000) 5:153-158; and Kim et al., *Anal. Biochem.* (2000) 284:42-48.

The NS3, NS4, NS5a, and NS5b polypeptides present in the various fusions described above can either be full-length polypeptides or portions of NS3, NS4 (NS4a and/or NS4b), NS5a, and NS5b polypeptides. The portions of NS3, NS4, NS5a, and NS5b polypeptides making up the fusion protein preferably comprise at least one epitope, which is recognized by a T cell receptor on an activated T cell, such as 2152-HEYPVGSQL-2160 (SEQ ID NO:1) and/or 2224-AELIEANLLWRQEMG-2238 (SEQ ID NO:2). Epitopes of NS3, NS4 (NS4a and NS4b), NS5a, NS5b, NS3NS4NS5a, and NS3NS4NS5aNS5b can be identified by several methods. For example, NS3, NS4, NS5a, NS5b polypeptides or fusion proteins comprising any combination of the above, can be isolated, for example, by immunoaffinity purification using a monoclonal antibody for the polypeptide or protein. The isolated protein sequence can then be screened by preparing a series of short peptides by proteolytic cleavage of the purified protein, which together span the entire protein sequence. By starting with, for example, 100-mer polypeptides, each polypeptide can be tested for the presence of epitopes recognized by a T-cell receptor on an HCV-activated T cell, progressively smaller and overlapping fragments can then be tested from an identified 100-mer to map the epitope of interest.

Epitopes recognized by a T-cell receptor on an HCV-activated T cell can be identified by, for example, ^{51}Cr release assay or by lymphoproliferation assay (see the examples). In a ^{51}Cr release assay, target cells can be constructed that display the epitope of interest by cloning a polynucleotide encoding the epitope into an expression vector and transforming the expression vector into the target cells. HCV-specific $CD8^+$ T cells will lyse target cells displaying, for example, an NS3, NS4, NS5a, NS5b, NS3NS4NS5a, or NS3NS4NS5aNS5b epitope and will not lyse cells

that do not display such an epitope. In a lymphoproliferation assay, HCV-activated $CD4^+$ T cells will proliferate when cultured with, for example, an NS3, NS4, NS5a, NS5b, NS3NS4NS5a, or NS3NS4NS5aNS5b epitopic peptide, but not in the absence of an HCV epitopic peptide.

NS3, NS4, NS5a, and NS5b polypeptides can occur in any order in the fusion protein. If desired, at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more of one or more of the polypeptides may occur in the fusion protein. Multiple viral strains of HCV occur, and NS3, NS4, NS5a, and NS5b polypeptides of any of these strains can be used in a fusion protein. A representative fusion protein for use in the present invention is shown if Figures 5A-5J. The depicted sequence includes amino acids 1242-3011 of the HCV polyprotein (representing polypeptides from NS3, NS4, NS5a and NS5b) with amino acids 1-121 of the HCV polyprotein (representing a polypeptide from the core region) fused to the C-terminus of NS5b. This numbering is relative to the HCV-1 polyprotein.

Nucleic acid and amino acid sequences of a number of HCV strains and isolates, including nucleic acid and amino acid sequences of NS3, NS4, NS5a, NS5b genes and polypeptides have been determined. For example, isolate HCV J1.1 is described in Kubo *et al.* (1989) Japan. Nucl. Acids Res. 17:10367-10372; Takeuchi *et al.*(1990) Gene 91:287-291; Takeuchi *et al.* (1990) J. Gen. Virol. 71:3027-3033; and Takeuchi *et al.* (1990) Nucl. Acids Res. 18:4626. The complete coding sequences of two independent isolates, HCV-J and BK, are described by Kato *et al.*, (1990) Proc. Natl. Acad. Sci. USA 87:9524-9528 and Takamizawa *et al.*, (1991) J. Virol. 65:1105-1113 respectively.

Publications that describe HCV-1 isolates include Choo *et al.* (1990) Brit. Med. Bull. 46:423-441; Choo *et al.* (1991) Proc. Natl. Acad. Sci. USA 88:2451-2455 and Han *et al.* (1991) Proc. Natl. Acad. Sci. USA 88:1711-1715. HCV isolates HC-J1 and HC-J4 are described in Okamoto *et al.* (1991) Japan J. Exp. Med. 60:167-177. HCV isolates HCT 18~, HCT 23, Th, HCT 27, EC1 and EC10 are described in Weiner *et al.* (1991) Virol. 180:842-848. HCV isolates Pt-1, HCV-K1 and HCV-K2 are described in Enomoto *et al.* (1990) Biochem. Biophys. Res.

Commun. 170:1021-1025. HCV isolates A, C, D & E are described in Tsukiyama-Kohara *et al.* (1991) Virus Genes 5:243-254.

Each of the NS3, NS4, NS5a, and NS5b components of a fusion protein can be obtained from the same HCV strain or isolate or from different HCV strains or isolates. Fusion proteins comprising HCV polypeptides from, for example, the NS3 polypeptide can be derived from a first strain of HCV, and the NS4, and NS5a polypeptides can be derived from a second strain of HCV. Alternatively, the NS4 polypeptide can be derived from a first strain of HCV, and the NS3 and NS5a polypeptides can be derived from a second strain of HCV. Optionally, the NS5a polypeptide can be derived from a first strain of HCV, and the NS3 and NS4 polypeptides can be derived from a second strain of HCV. NS3, NS4 and NS5a polypeptides that are each derived from different HCV strains can also be used in an HCV fusion protein. Similarly, in a fusion protein comprising NS5b, at least one of the NS3, NS4, NS5a, and NS5b polypeptides can be derived from a different HCV strain than the other polypeptides. Optionally, NS3, NS4, NS5a, and NS5b polypeptides that are each derived from different HCV strains can also be used in an NS3NS4NS5aNS5b fusion protein.

In addition to NS3, NS4a, NS4b, NS5a and NS5b, the fusion proteins can contain other polypeptides derived from the HCV polyprotein. For example, it may be desirable to include polypeptides derived from the core region of the HCV polyprotein. This region occurs at amino acid positions 1-191 of the HCV polyprotein, numbered relative to HCV-1. Either the full-length protein, fragments thereof, such as amino acids 1-150, e.g., amino acids 1-130, 1-120, for example, amino acids 1-121, 1-122, 1-123, etc., or smaller fragments containing epitopes of the full-length protein may be used in the subject fusions, such as those epitopes found between amino acids 10-53, amino acids 10-45, amino acids 67-88, amino acids 120-130, or any of the core epitopes identified in, e.g., Houghton et al., U.S. Patent No. 5,350,671; Chien et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:10011-10015; Chien et al., *J. Gastroent. Hepatol.* (1993) 8:S33-39; Chien et al., International Publication No. WO 93/00365; Chien, D.Y., International Publication No. WO 94/01778; and U.S. Patent Nos. 6,280,927 and 6,150,087. Moreover, a

protein resulting from a frameshift in the core region of the polyprotein, such as described in International Publication No. WO 99/63941, may be used. The fusions may also contain polynucleotides encoding E1E2 polypeptides, as described further below.

Preferably, the above-described fusion proteins, as well as the individual components of these proteins, are produced recombinantly. A polynucleotide encoding these proteins can be introduced into an expression vector which can be expressed in a suitable expression system. A variety of bacterial, yeast, mammalian and insect expression systems are available in the art and any such expression system can be used. Optionally, a polynucleotide encoding these proteins can be translated in a cell-free translation system. Such methods are well known in the art. The proteins also can be constructed by solid phase protein synthesis.

If desired, the fusion proteins, or the individual components of these proteins, also can contain other non-HCV amino acid sequences, such as amino acid linkers or signal sequences, as well as ligands useful in protein purification, such as glutathione-S-transferase and staphylococcal protein A.

E1E2 Polypeptides

As explained above, the compositions of the present invention may also include E1 and E2 polypeptides, complexes of these polypeptides or polynucleotides encoding the same. The E1 and E2 polypeptides and complexes thereof can be provided independent of the nonstructural fusion protein or can be incorporated into the same fusion. Moreover, E1E2 complexes can be provided as proteins, or as polynucleotides encoding the same.

In this regard, E1, E2 and p7 are known to contain human T-cell epitopes (both CD4+ and CD8+) and including one or more of these epitopes serves to increase vaccine efficacy as well as to increase protective levels against multiple HCV genotypes. Moreover, multiple copies of specific, conserved T-cell epitopes can also be used in E1E2 complexes, such as a composite of epitopes from different genotypes.

As explained above, the E1 and E2 polypeptides that make up the E1E2 complexes can be associated either through non-covalent or covalent interactions. Such complexes may be made up of immunogenic fragments of E1 and E2 which comprise epitopes. For example, fragments of E1 polypeptides can comprise from about 5 to nearly the full-length of the molecule, such as 6, 10, 25, 50, 75, 100, 125, 150, 175, 185 or more amino acids of an E1 polypeptide, or any integer between the stated numbers. Similarly, fragments of E2 polypeptides can comprise 6, 10, 25, 50, 75, 100, 150, 200, 250, 300, or 350 amino acids of an E2 polypeptide, or any integer between the stated numbers. The E1 and E2 polypeptides may be from the same or different HCV strains. For example, epitopes derived from, e.g., the hypervariable region of E2, such as a region spanning amino acids 384-410 or 390-410, can be included in the E2 polypeptide. A particularly effective E2 epitope to incorporate into the E2 sequence or E1E2 complexes is one which includes a consensus sequence derived from this region, such as the consensus sequence Gly-Ser-Ala-Ala-Arg-Thr-Thr-Ser-Gly-Phe-Val-Ser-Leu-Phe-Ala-Pro-Gly-Ala-Lys-Gln-Asn (SEQ ID NO:5), which represents a consensus sequence for amino acids 390-410 of the HCV type 1 genome. Additional epitopes of E1 and E2 are known and described in, e.g., Chien et al., International Publication No. WO 93/00365.

Moreover, the E1 and E2 polypeptides may lack all or a portion of the membrane spanning domain. The membrane anchor sequence functions to associate the polypeptide to the endoplasmic reticulum. Normally, such polypeptides are capable of secretion into growth medium in which an organism expressing the protein is cultured. However, as described in International Publication No. WO 98/50556, such polypeptides may also be recovered intracellularly. Secretion into growth medium is readily determined using a number of detection techniques, including, e.g., polyacrylamide gel electrophoresis and the like, and immunological techniques such as immunoprecipitation assays as described in, e.g., International Publication No. WO 96/04301, published February 15, 1996. With E1, generally polypeptides terminating with about amino acid position 370 and higher (based on the numbering of HCV1 E1) will be retained by the ER and hence not secreted into growth media. With E2, polypeptides terminating with about amino acid position

731 and higher (also based on the numbering of the HCV1 E2 sequence) will be retained by the ER and not secreted. (See, e.g., International Publication No. WO 96/04301, published February 15, 1996). It should be noted that these amino acid positions are not absolute and may vary to some degree. Thus, the present invention contemplates the use of E1 and E2 polypeptides which retain the transmembrane binding domain, as well as polypeptides which lack all or a portion of the transmembrane binding domain, including E1 polypeptides terminating at about amino acids 369 and lower, and E2 polypeptides, terminating at about amino acids 730 and lower, are intended to be captured by the present invention. Furthermore, the C-terminal truncation can extend beyond the transmembrane spanning domain towards the N-terminus. Thus, for example, E1 truncations occurring at positions lower than, e.g., 360 and E2 truncations occurring at positions lower than, e.g., 715, are also encompassed by the present invention. All that is necessary is that the truncated E1 and E2 polypeptides remain functional for their intended purpose. However, particularly preferred truncated E1 constructs are those that do not extend beyond about amino acid 300. Most preferred are those terminating at position 360. Preferred truncated E2 constructs are those with C-terminal truncations that do not extend beyond about amino acid position 715. Particularly preferred E2 truncations are those molecules truncated after any of amino acids 715-730, such as 725. If truncated molecules are used, it is preferable to use E1 and E2 molecules that are both truncated.

E2 exists as multiple species (Spaete et al., *Virol.* (1992) 188:819-830; Selby et al., *J. Virol.* (1996) 70:5177-5182; Grakoui et al., *J. Virol.* (1993) 67:1385-1395; Tomei et al., *J. Virol.* (1993) 67:4017-4026) and clipping and proteolysis may occur at the N– and C-termini of the E1 and E2 polypeptides. Thus, an E2 polypeptide for use herein may comprise at least amino acids 405-661, e.g., 400, 401, 402... to 661, such as 384-661, 384-715, 384-746, 384-749 or 384-809, or 384 to any C-terminus between 661-809, of an HCV polyprotein, numbered relative to the full-length HCV-1 polyprotein. Similarly, preferable E1 polypeptides for use herein can comprise amino acids 192-326, 192-330, 192-333, 192-360, 192-363, 192-383, or 192 to any C-terminus between 326-383, of an HCV polyprotein.

The E1 and E2 polypeptides and complexes thereof may also be present as asialoglycoproteins. Such asialoglycoproteins are produced by methods known in the art, such as by using cells in which terminal glycosylation is blocked. When these proteins are expressed in such cells and isolated by GNA lectin affinity chromatography, the E1 and E2 proteins aggregate spontaneously. Detailed methods for producing these E1E2 aggregates are described in, e.g., U.S. Patent No. 6,074,852. For example, E1E2 complexes are readily produced recombinantly, either as fusion proteins or by e.g., co-transfecting host cells with constructs encoding for the E1 and E2 polypeptides of interest. Co-transfection can be accomplished either in *trans* or *cis*, i.e., by using separate vectors or by using a single vector which bears both of the E1 and E2 genes. If done using a single vector, both genes can be driven by a single set of control elements or, alternatively, the genes can be present on the vector in individual expression cassettes, driven by individual control elements. Following expression, the E1 and E2 proteins will spontaneously associate. Alternatively, the complexes can be formed by mixing the individual proteins together which have been produced separately, either in purified or semi-purified form, or even by mixing culture media in which host cells expressing the proteins, have been cultured, if the proteins are secreted. Finally, the E1E2 complexes of the present invention may be expressed as a fusion protein wherein the desired portion of E1 is fused to the desired portion of E2.

Moreover, the E1E2 complexes may be present as a heterogeneous mixture of molecules, due to clipping and proteolytic cleavage, as described above. Thus, a composition including E1E2 complexes may include multiple species of E1E2, such as E1E2 terminating at amino acid 746 ($E1E2_{746}$), E1E2 terminating at amino acid 809 ($E1E2_{809}$), or any of the other various E1 and E2 molecules described above, such as E2 molecules with N-terminal truncations of from 1-20 amino acids, such as E2 species beginning at amino acid 387, amino acid 402, amino acid 403, etc.

E1E2 complexes are readily produced recombinantly, either as fusion proteins or by e.g., co-transfecting host cells with constructs encoding for the E1 and E2 polypeptides of interest. Co-transfection can be accomplished either in *trans* or *cis*, i.e., by using separate vectors or by using a single vector which bears both of the E1

and E2 genes. If done using a single vector, both genes can be driven by a single set of control elements or, alternatively, the genes can be present on the vector in individual expression cassettes, driven by individual control elements. Following expression, the E1 and E2 proteins will spontaneously associate. Alternatively, the complexes can be formed by mixing the individual proteins together which have been produced separately, either in purified or semi-purified form, or even by mixing culture media in which host cells expressing the proteins, have been cultured, if the proteins are secreted. Finally, the E1E2 complexes of the present invention may be expressed as a fusion protein wherein the desired portion of E1 is fused to the desired portion of E2.

Methods for producing E1E2 complexes from full-length, truncated E1 and E2 proteins which are secreted into media, as well as intracellularly produced truncated proteins, are known in the art. For example, such complexes may be produced recombinantly, as described in U.S. Patent No. 6,121,020; Ralston et al., *J. Virol.* (1993) 67:6753-6761, Grakoui et al., *J. Virol.* (1993) 67:1385-1395; and Lanford et al., *Virology* (1993) 197:225-235.

Polynucleotides Encoding the Fusion Proteins and E1E2 Complexes

Polynucleotides contain less than an entire HCV genome and can be RNA or single- or double-stranded DNA. Preferably, the polynucleotides are isolated free of other components, such as proteins and lipids. The polynucleotides encode the fusion proteins, E1 and E2 polypeptides and complexes thereof, described above, and thus comprise coding sequences thereof. Polynucleotides of the invention can also comprise other non-HCV nucleotide sequences, such as sequences coding for linkers, signal sequences, or ligands useful in protein purification such as glutathione-S-transferase and staphylococcal protein A.

Polynucleotides encoding the various HCV polypeptides can be isolated from a genomic library derived from nucleic acid sequences present in, for example, the plasma, serum, or liver homogenate of an HCV infected individual or can be synthesized in the laboratory, for example, using an automatic synthesizer. An

amplification method such as PCR can be used to amplify polynucleotides from either HCV genomic DNA or cDNA encoding therefor.

Polynucleotides can comprise coding sequences for these polypeptides which occur naturally or can include artificial sequences which do not occur in nature. These polynucleotides can be ligated to form a coding sequence for the fusion proteins and E1E2 complexes using standard molecular biology techniques. If desired, polynucleotides can be cloned into an expression vector and transformed into, for example, bacterial, yeast, insect, or mammalian cells so that the fusion proteins of the invention can be expressed in and isolated from a cell culture.

The expression constructs of the present invention, including the desired fusion, or individual expression constructs comprising the individual components of these fusions, may be used for nucleic acid immunization, to stimulate an immunological response, such as a cellular immune response, using standard gene delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject. For example, the constructs can be delivered as plasmid DNA, e.g., contained within a plasmid, such as pBR322, pUC, or ColE1

Additionally, the expression constructs can be packaged in liposomes prior to delivery to the cells. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use with the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA)

liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes. The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in METHODS OF IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; Papahadjopoulos et al., *Biochim. Biophys. Acta* (1975) 394:483; Wilson et al., *Cell* (1979) 17:77); Deamer and Bangham, *Biochim. Biophys. Acta* (1976) 443:629; Ostro et al., *Biochem. Biophys. Res. Commun.* (1977) 76:836; Fraley et al., *Proc. Natl. Acad. Sci. USA* (1979) 76:3348); Enoch and Strittmatter, *Proc. Natl. Acad. Sci. USA* (1979) 76:145); Fraley et al., *J. Biol. Chem.* (1980) 255:10431; Szoka and Papahadjopoulos, *Proc. Natl. Acad. Sci. USA* (1978) 75:145; and Schaefer-Ridder et al., *Science* (1982) 215:166.

The DNA can also be delivered in cochleate lipid compositions similar to those described by Papahadjopoulos et al., *Biochem. Biophys. Acta.* (1975) 394:483-491. See, also, U.S. Patent Nos. 4,663,161 and 4,871,488.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems, such as murine sarcoma virus, mouse mammary tumor virus, Moloney murine leukemia virus, and Rous sarcoma virus. A selected gene can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109. Briefly,

retroviral gene delivery vehicles of the present invention may be readily constructed from a wide variety of retroviruses, including for example, B, C, and D type retroviruses as well as spumaviruses and lentiviruses such as FIV, HIV, HIV-1, HIV-2 and SIV (see RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985). Such retroviruses may be readily obtained from depositories or collections such as the American Type Culture Collection ("ATCC"; 10801 University Blvd., Manassas, VA 20110-2209), or isolated from known sources using commonly available techniques.

A number of adenovirus vectors have also been described, such as adenovirus Type 2 and Type 5 vectors. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as but not limited to vectors derived from the Sindbis and Semliki Forest viruses, VEE, will also find use as viral vectors for delivering the gene of interest. For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072.

Other vectors can be used, including but not limited to simian virus 40 and cytomegalovirus. Bacterial vectors, such as Salmonella ssp. *Yersinia enterocolitica*, Shigella spp., *Vibrio cholerae*, Mycobacterium strain BCG, and *Listeria monocytogenes* can be used. Minichromosomes such as MC and MC1, bacteriophages, cosmids (plasmids into which phage lambda *cos* sites have been inserted) and replicons (genetic elements that are capable of replication under their own control in a cell) can also be used.

The expression constructs may also be encapsulated, adsorbed to, or associated with, particulate carriers. Such carriers present multiple copies of a selected molecule to the immune system and promote trapping and retention of molecules in local lymph nodes. The particles can be phagocytosed by macrophages and can enhance antigen presentation through cytokine release. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; and McGee et al., *J. Microencap.* (1996).

One preferred method for adsorbing macromolecules onto prepared microparticles is described in International Publication No. WO 00/050006. Briefly, microparticles are rehydrated and dispersed to an essentially monomeric suspension of microparticles using dialyzable anionic or cationic detergents. Useful detergents include, but are not limited to, any of the various N-methylglucamides (known as MEGAs), such as heptanoyl-N-methylglucamide (MEGA-7), octanoyl-N-methylglucamide (MEGA-8), nonanoyl-N-methylglucamide (MEGA-9), and decanoyl-N-methyl-glucamide (MEGA-10); cholic acid; sodium cholate; deoxycholic acid; sodium deoxycholate; taurocholic acid; sodium taurocholate; taurodeoxycholic acid; sodium taurodeoxycholate; 3-[(3-cholamidopropyl)dimethylammonio] -1-propane-sulfonate (CHAPS); 3-[(3-cholamidopropyl) dimethylammonio]-2-hydroxy-1-propane-sulfonate (CHAPSO); -dodecyl-N,N-dimethyl-3-ammonio-1-propane-sulfonate (ZWITTERGENT 3-12); N,N-bis-(3-D-gluconeamidopropyl)-deoxycholamide (DEOXY-BIGCHAP); -octylglucoside; sucrose monolaurate; glycocholic acid/sodium glycocholate; laurosarcosine (sodium salt); glycodeoxycholic acid/sodium glycodeoxycholate; sodium dodceyl sulfate (SDS); 3-(trimethylsilyl)-1-propanesulfonic acid (DSS); cetrimide (CTAB, the principal component of which is hexadecyltrimethylammonium bromide); hexadecyltrimethylammonium bromide; dodecyltrimethylammonium bromide; hexadecyltrimethyl-ammonium bromide; tetradecyltrimethylammonium bromide; benzyl dimethyldodecylammonium bromide; benzyl dimethyl-hexadecylammonium chloride; and benzyl dimethyltetra-

decylammonium bromide. The above detergents are commercially available from e.g., Sigma Chemical Co., St. Louis, MO. Various cationic lipids known in the art can also be used as detergents. See Balasubramaniam et al., 1996, *Gene Ther.*, 3:163-72 and Gao, X., and L. Huang. 1995, *Gene Ther.*, 2:7110-722.

A wide variety of other methods can be used to deliver the expression constructs to cells. Such methods include DEAE dextran-mediated transfection, calcium phosphate precipitation, polylysine- or polyornithine-mediated transfection, or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like. Other useful methods of transfection include electroporation, sonoporation, protoplast fusion, liposomes, peptoid delivery, or microinjection. See, e.g., Sambrook et al., *supra*, for a discussion of techniques for transforming cells of interest; and Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Methods of delivering DNA using electroporation are described in, e.g., U.S. Patent Nos. 6,132,419; 6,451,002, 6,418,341, 6233,483, U.S. Patent Publication No. 2002/0146831; and International Publication No. WO/0045823.

Moreover, the HCV polynucleotides can be adsorbed to, or entrapped within, an ISCOM. Classic ISCOMs are formed by combination of cholesterol, saponin, phospholipid, and immunogens, such as viral envelope proteins. Generally, the HCV molecules (usually with a hydrophobic region) are solubilized in detergent and added to the reaction mixture, whereby ISCOMs are formed with the HCV molecule incorporated therein. ISCOM matrix compositions are formed identically, but without viral proteins. Proteins with high positive charge may be electrostatically bound in the ISCOM particles, rather than through hydrophobic forces. For a more detailed general discussion of saponins and ISCOMs, and methods of formulating ISCOMs, see Barr et al. (1998) *Adv. Drug Delivery Reviews* 32:247-271 (1998); U.S. Patent Nos. 4,981,684, 5,178,860, 5,679,354 and 6,027,732; European Publ. Nos. EPA 109,942; 180,564 and 231,039; and Coulter et al. (1998) *Vaccine* 16:1243.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering the expression constructs of

the present invention. The particles are coated with the construct to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744.

Compositions Comprising Fusion Proteins or Polynucleotides

The invention also provides compositions comprising the fusion proteins or polynucleotides, as well as compositions including the individual components of these fusion proteins or polynucleotides. Compositions of the invention preferably comprise a pharmaceutically acceptable carrier. The carrier should not itself induce the production of antibodies harmful to the host. Pharmaceutically acceptable carriers are well known to those in the art. Such carriers include, but are not limited to, large, slowly metabolized, macromolecules, such as proteins, polysaccharides such as latex functionalized sepharose, agarose, cellulose, cellulose beads and the like, polylactic acids, polyglycolic acids, polymeric amino acids such as polyglutamic acid, polylysine, and the like, amino acid copolymers, and inactive virus particles.

Pharmaceutically acceptable salts can also be used in compositions of the invention, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as salts of organic acids such as acetates, proprionates, malonates, or benzoates. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those of skill in the art. Compositions of the invention can also contain liquids or excipients, such as water, saline, glycerol, dextrose, ethanol, or the like, singly or in combination, as well as substances such as wetting agents, emulsifying agents, or pH buffering agents. Liposomes can also be used as a carrier for a composition of the invention, such liposomes are described above.

If desired, co-stimulatory molecules which improve immunogen presentation to lymphocytes, such as B7-1 or B7-2, or cytokines such as GM-CSF, IL-2, and IL-

12, can be included in a composition of the invention. Optionally, adjuvants can also be included in a composition. Adjuvants which can be used include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (U.S. Patent No. 6,299,884; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% TWEEN 80™, and 0.5% SPAN 85™ (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% TWEEN 80™, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) RIBI™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% TWEEN 80™, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DETOX™); (3) saponin adjuvants, such as QS21 or STIMULON™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), which ISCOMs may be devoid of additional detergent, see, e.g., International Publication No. WO 00/07621; (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 (International Publication No. WO 99/44636), etc.), interferons (e.g., gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63) LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino

acid at position 9 and glycine substituted at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); (7) MPL or 3-O-deacylated MPL (3dMPL) (see, e.g., GB 2220221), EP-A-0689454, optionally in the substantial absence of alum when used with pneumococcal saccharides (see, e.g., International Publication No. WO 00/56358); (8) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions (see, e.g., EP-A-0835318, EP-A-0735898, EP-A-0761231; (9) oligonucleotides comprising CpG motifs (see, e.g., Roman et al. (1997) *Nat. Med.* 3:849-854; Weiner et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:10833-10837; Davis et al. (1998) *J. Immunol.* 160:870-876; Chu et al. (1997) *J. Exp. Med.* 186:1623-1631; Lipford et al. (1997) *Eur. J. Immunol.* 27:2340-2344; Moldoveanu et al. (1988) *Vaccine* 16:1216-1224; Krieg et al. (1995) *Nature* 374:546-549; Klinman et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:2879-2883; Ballas et al. (1996) *J. Immunol.* 157:1840-1845; Cowdery et al. (1996) *J. Immunol.* 156:4570-4575; Halpern et al. (1996) *Cell Immunol.* 167:72-78; Yamamoto et al. (1988) *Jpn. J. Cancer Res.* 79:866-873; Stacey et al. (1996) *J. Immunol.* 157:2116-2122; Messina et al. (1991) *J. Immunol.* 147:1759-1764; Yi et al. (1996) *J. Immunol.* 157:4918-4925; Yi et al. (1996) *J. Immunol.* 157:5394-5402; Yi et al. (1998) *J. Immunol.* 160:4755-4761; Yi et al. (1998) *J. Immunol.* 160:5898-5906; International Publication Nos. WO 96/02555, WO 98/16247, WO 98/18810, WO 98/40100, WO 98/55495, WO 98/37919 and WO 98/52581), such as those containing at least on CG dinucleotide, with cytosine optionally replaced with 5-methylcytosine; (10) a polyoxyethylene ether or a polyoxyethylene ester (see, e.g., International Publication No. WO 99/52549); (11) a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol (see, e.g., International Publication No. WO 01/21207) or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol (see, e.g., International Publication No. WO 01/21152); (12) a saponin and an immunostimulatory oligonucleotide such as a CpG oligonucleotide (see, e.g., International Publication No. WO 00/62800); (13) an immunostimulant and a particle of metal salt (see, e.g., International Publication No. WO 00/23105); and (14) other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), -acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(l'-2'-dipalmitoyl-*sn*-glycero-3-huydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Moreover, the HCV proteins can be adsorbed to, or entrapped within, an ISCOM, as described above. Additionally, ISCOMs with adsorbed HCV core proteins, either the entire core region or a fragment of HCV core protein, may be added to the formulations. Most preferably, the HCV core protein is a fragment comprising a polypeptide from the region spanning amino acid positions 121-135. See, e.g., International Publication No. WO 01/37869A.

As explained above, the composition may also contain immunostimulatory molecules, either in addition to or in place of the antigen delivery system. Immunostimulatory agents for use herein include, without limitation, monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™). MPL may be formulated into an emulsion to enhance its immunostimulatory affect. See, e.g., Ulrich et al., "MPLr immunostimulat: adjuvant formulations." in Vaccine Adjuvants: Prepartion Methods and Research Protocols (O'Hagan DT, ed.) Human Press Inc., NJ (2000) pp. 273-282. MPL has been shown to induce the synthesis and release of cytokines, particularly IL-2 and IFN-γ. Other useful immunostimulatory molecules include LPS and immunostimulatory nucleic acid sequences (ISS), including but not limited to, unmethylated CpG motifs, such as CpG oligonucleotides.

Oligonucleotides containing unmethylated CpG motifs have been shown to induce activation of B cells, NK cells and antigen-presenting cells (APCs), such as monocytes and macrophages. See, e.g., U.S. Patent No. 6,207,646. Thus, adjuvants derived from the CpG family of molecules, CpG dinucleotides and synthetic oligonucleotides which comprise CpG motifs (see, e.g., Krieg et al. *Nature* (1995) 374:546 and Davis et al. *J. Immunol.* (1998) 160:870-876) such as any of the various immunostimulatory CpG oligonucleotides disclosed in U.S. Patent No. 6,207,646, may be used in the subject methods and compositions. Such CpG oligonucleotides generally comprise at least 8 up to about 100 basepairs, preferably 8 to 40 basepairs,

more preferably 15-35 basepairs, preferably 15-25 basepairs, and any number of basepairs between these values. For example, oligonucleotides comprising the consensus CpG motif, represented by the formula 5'-X_1CGX_2-3', where X_1 and X_2 are nucleotides and C is unmethylated, will find use as immunostimulatory CpG molecules. Generally, X_1 is A, G or T, and X_2 is C or T. Other useful CpG molecules include those captured by the formula 5'-X_1X_2CGX_3X_4, where X_1 and X_2 are a sequence such as GpT, GpG, GpA, ApA, ApT, ApG, CpT, CpA, CpG, TpA, TpT or TpG, and X_3 and X_4 are TpT, CpT, ApT, ApG, CpG, TpC, ApC, CpC, TpA, ApA, GpT, CpA, or TpG, wherein "p" signifies a phosphate bond. Preferably, the oligonucleotides do not include a GCG sequence at or near the 5'- and/or 3' terminus. Additionally, the CpG is preferably flanked on its 5'-end with two purines (preferably a GpA dinucleotide) or with a purine and a pyrimidine (preferably, GpT), and flanked on its 3'-end with two pyrimidines, preferably a TpT or TpC dinucleotide. Thus, preferred molecules will comprise the sequence GACGTT, GACGTC, GTCGTT or GTCGCT, and these sequences will be flanked by several additional nucleotides. The nucleotides outside of this central core area appear to be extremely amendable to change.

Moreover, the CpG oligonucleotides for use herein may be double- or single-stranded. Double-stranded molecules are more stable *in vivo* while single-stranded molecules display enhanced immune activity. Additionally, the phosphate backbone may be modified, such as phosphorodithioate-modified, in order to enhance the immunostimulatory activity of the CpG molecule. As described in U.S. Patent No. 6,207,646, CpG molecules with phosphorothioate backbones preferentially activate B-cells, while those having phosphodiester backbones preferentially activate monocytic (macrophages, dendritic cells and monocytes) and NK cells.

One exemplary CpG oligonucleotide for use in the present compositions has the sequence 5'-TCCATGACGTTCCTGACGTT-3' (SEQ ID NO:6).

CpG molecules can readily be tested for their ability to stimulate an immune response using standard techniques, well known in the art. For example, the ability of the molecule to stimulate a humoral and/or cellular immune response is readily determined using the immunoassays described above. Moreover, the antigen and

adjuvant compositions can be administered with and without the CpG molecule to determine whether an immune response is enhanced.

The HCV proteins may also be encapsulated, adsorbed to, or associated with, particulate carriers, as described above with reference to the HCV polynucleotides. As explained above, examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; and McGee et al., *J. Microencap.* (1996). One preferred method for adsorbing macromolecules onto prepared microparticles is described above and in International Publication No. WO 00/050006.

Methods of Producing HCV-Specific Antibodies

The HCV proteins can be used to produce HCV-specific polyclonal and monoclonal antibodies. HCV-specific polyclonal and monoclonal antibodies specifically bind to HCV antigens. Polyclonal antibodies can be produced by administering the fusion protein to a mammal, such as a mouse, a rabbit, a goat, or a horse. Serum from the immunized animal is collected and the antibodies are purified from the plasma by, for example, precipitation with ammonium sulfate, followed by chromatography, preferably affinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art.

Monoclonal antibodies directed against HCV-specific epitopes present in the proteins can also be readily produced. Normal B cells from a mammal, such as a mouse, immunized with an HCV protein, can be fused with, for example, HAT-sensitive mouse myeloma cells to produce hybridomas. Hybridomas producing HCV-specific antibodies can be identified using RIA or ELISA and isolated by cloning in semi-solid agar or by limiting dilution. Clones producing HCV-specific antibodies are isolated by another round of screening.

Antibodies, either monoclonal and polyclonal, which are directed against HCV epitopes, are particularly useful for detecting the presence of HCV or HCV antigens in a sample, such as a serum sample from an HCV-infected human. An immunoassay for an HCV antigen may utilize one antibody or several antibodies. An

immunoassay for an HCV antigen may use, for example, a monoclonal antibody directed towards an HCV epitope, a combination of monoclonal antibodies directed towards epitopes of one HCV polypeptide, monoclonal antibodies directed towards epitopes of different HCV polypeptides, polyclonal antibodies directed towards the same HCV antigen, polyclonal antibodies directed towards different HCV antigens, or a combination of monoclonal and polyclonal antibodies. Immunoassay protocols may be based, for example, upon competition, direct reaction, or sandwich type assays using, for example, labeled antibody. The labels may be, for example, fluorescent, chemiluminescent, or radioactive.

The polyclonal or monoclonal antibodies may further be used to isolate HCV particles or antigens by immunoaffinity columns. The antibodies can be affixed to a solid support by, for example, adsorption or by covalent linkage so that the antibodies retain their immunoselective activity. Optionally, spacer groups may be included so that the antigen binding site of the antibody remains accessible. The immobilized antibodies can then be used to bind HCV particles or antigens from a biological sample, such as blood or plasma. The bound HCV particles or antigens are recovered from the column matrix by, for example, a change in pH.

HCV-Specific T cells

HCV-specific T cells that are activated by the above-described fusions and E1E2 complexes, including the NS3NS4NS5a fusion protein or NS3NS4NS5aNS5b fusion protein, and the E1E2 complexes, expressed *in vivo* or *in vitro*, preferably recognize an epitope of an HCV polypeptide such as an E1, E2, NS3, NS4, NS5a, NS5b polypeptide, including an epitope of an NS3NS4NS5a fusion protein or an NS3NS4NS5aNS5b fusion protein, or an E1E2 complex. HCV-specific T cells can be $CD8^+$ or $CD4^+$.

HCV-specific $CD8^+$ T cells preferably are cytotoxic T lymphocytes (CTL) which can kill HCV-infected cells that display E1, E2, NS3, NS4, NS5a, NS5b epitopes complexed with an MHC class I molecule. HCV-specific $CD8^+$ T cells may also express interferon-γ (IFN-γ). HCV-specific $CD8^+$ T cells can be detected by, for example, ^{51}Cr release assays (see the examples). ^{51}Cr release assays measure the

ability of HCV-specific $CD8^+$ T cells to lyse target cells displaying an E1, E2, E1E2, NS3, NS4, NS5a, NS5b, NS3NS4NS5a, or NS3NS4NS5aNS5b epitope. HCV-specific $CD8^+$ T cells which express IFN-γ can also be detected by immunological methods, preferably by intracellular staining for IFN-γ after *in vitro* stimulation with an E1, E2, NS3, an NS4, an NS5a, or an NS5b polypeptide (see the examples).

HCV-specific $CD4^+$ cells activated by the above-described E1E2 complexes and fusions, such as an E1 polypeptide, an E2 polypeptide, an E1E2 complex, NS3NS4NS5a or NS3NS4NS5aNS5b fusion protein, expressed *in vivo* or *in vitro*, preferably recognize an epitope of an E1, E2, NS3, NS4, NS5a, or NS5b polypeptide, including an epitope of an E1E2 complex, NS3NS4NS5a or NS3NS4NS5aNS5b fusion protein, that is bound to an MHC class II molecule on an HCV-infected cell and proliferate in response to stimulating E1E2 complexes with NS3NS4NS5a or NS3NS4NS5aNS5b peptides, with or without a core polypeptide.

HCV-specific $CD4^+$ T cells can be detected by a lymphoproliferation assay (see examples). Lymphoproliferation assays measure the ability of HCV-specific $CD4^+$ T cells to proliferate in response to an E1, E2, NS3, an NS4, an NS5a, or an NS5b epitope.

Methods of Activating HCV-Specific T Cells.

The HCV proteins or polynucleotides can be used to stimulate an immune response, such as to activate HCV-specific T cells either *in vitro* or *in vivo*. Activation of HCV-specific T cells can be used, *inter alia*, to provide model systems to optimize CTL responses to HCV and to provide prophylactic or therapeutic treatment against HCV infection. For *in vitro* activation, proteins are preferably supplied to T cells via a plasmid or a viral vector, such as an adenovirus vector, as described above.

Polyclonal populations of T cells can be derived from the blood, and preferably
from peripheral lymphoid organs, such as lymph nodes, spleen, or thymus, of mammals that have been infected with an HCV. Preferred mammals include mice, chimpanzees, baboons, and humans. The HCV serves to expand the number of

activated HCV-specific T cells in the mammal. The HCV-specific T cells derived from the mammal can then be restimulated *in vitro* by adding, e.g., HCV E1E2 and NS3NS4NS5a or NS3NS4NS5aNS5b epitopic peptides, with or without a core polypeptide, to the T cells. The HCV-specific T cells can then be tested for, *inter alia*, proliferation, the production of IFN-γ, and the ability to lyse target cells displaying E1E2, NS3NS4NS5a or NS3NS4NS5aNS5b epitopes *in vitro*.

In a lymphoproliferation assay (see examples), HCV-activated $CD4^+$ T cells proliferate when cultured with an NS3, NS4, NS5a, NS5b, NS3NS4NS5a, or NS3NS4NS5aNS5b epitopic peptide, but not in the absence of an epitopic peptide. Thus, particular E1, E2, NS3, NS4, NS5a, NS5b, NS3NS4NS5a and NS3NS4NS5aNS5b epitopes that are recognized by HCV-specific $CD4^+$ T cells can be identified using a lymphoproliferation assay.

Similarly, detection of IFN-γ in HCV-specific $CD8^+$ T cells after *in vitro* stimulation with the above-described HCV proteins, can be used to identify E1, E2, E1E2, NS3, NS4, NS5a, NS5b, NS3NS4NS5a, and NS3NS4NS5aNS5b epitopes that particularly effective at stimulating $CD8^+$ T cells to produce IFN-γ (see examples).

Further, ^{51}Cr release assays are useful for determining the level of CTL response to HCV. *See* Cooper *et al.* Immunity 10:439-449. For example, HCV-specific $CD8^+$ T cells can be derived from the liver of an HCV infected mammal. These T cells can be tested in ^{51}Cr release assays against target cells displaying, e.g., E1E2, NS3NS4NS5a and/or NS3NS4NS5aNS5b epitopes. Several target cell populations expressing different E1E2, NS3NS4NS5a and/or NS3NS4NS5aNS5b epitopes can be constructed so that each target cell population displays different epitopes of E1E2, NS3NS4NS5a and/or NS3NS4NS5aNS5b. The HCV-specific $CD8^+$ cells can be assayed against each of these target cell populations. The results of the ^{51}Cr release assays can be used to determine which epitopes of E1E2, NS3NS4NS5a and/or NS3NS4NS5aNS5b are responsible for the strongest CTL response to HCV. E1E2 complexes, NS3NS4NS5a fusion proteins or NS3NS4NS5aNS5b fusion proteins, with or without core polypeptides, which

contain the epitopes responsible for the strongest CTL response can then be constructed using the information derived from the ^{51}Cr release assays.

HCV proteins as described above, or polynucleotides encoding such proteins, can be administered to a mammal, such as a mouse, baboon, chimpanzee, or human, to stimulate an immune response, such as to activate HCV-specific T cells *in vivo*. Administration can be by any means known in the art, including parenteral, intranasal, intramuscular or subcutaneous injection, including injection using a biological ballistic gun ("gene gun"), as discussed above.

Preferably, injection of an HCV polynucleotide is used to activate T cells. In addition to the practical advantages of simplicity of construction and modification, injection of the polynucleotides results in the synthesis of a fusion protein in the host. Thus, these immunogens are presented to the host immune system with native post-translational modifications, structure, and conformation. The polynucleotides are preferably injected intramuscularly to a large mammal, such as a human, at a dose of 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg/kg.

A composition of the invention comprising the HCV proteins or polynucleotides is administered in a manner compatible with the particular composition used and in an amount which is effective to stimulate an immune response, such as to activate HCV-specific T cells as measured by, *inter alia*, a ^{51}Cr release assay, a lymphoproliferation assay, or by intracellular staining for IFN-γ. The proteins and/or polynucleotides can be administered either to a mammal which is not infected with an HCV or can be administered to an HCV-infected mammal. The particular dosages of the polynucleotides or proteins in a composition will depend on many factors including, but not limited to the species, age, and general condition of the mammal to which the composition is administered, and the mode of administration of the composition. An effective amount of the composition of the invention can be readily determined using only routine experimentation. *In vitro* and *in vivo* models described above can be employed to identify appropriate doses. The amount of polynucleotide used in the example described below provides general guidance which can be used to optimize the activation of HCV-specific T cells either

in vivo or *in vitro*. Generally, 0.5, 0.75, 1.0, 1.5, 2.0, 2.5, 5 or 10 mg of an HCV fusion and E1 and E2 polypeptides, such as an E1E2 complex, an NS3NS4NS5a or NS3NS4NS5aNS5b fusion protein or polynucleotide, with or without a core polypeptide, will be administered to a large mammal, such as a baboon, chimpanzee, or human. If desired, co-stimulatory molecules or adjuvants can also be provided before, after, or together with the compositions.

Immune responses of the mammal generated by the delivery of a composition of the invention, including activation of HCV-specific T cells, can be enhanced by varying the dosage, route of administration, or boosting regimens. Compositions of the invention may be given in a single dose schedule, or preferably in a multiple dose schedule in which a primary course of vaccination includes 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and/or reenforce an immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose or doses after several months.

Deposits of Strains Useful in Practicing the Invention

A deposit of biologically pure cultures of the following strains was made with the American Type Culture Collection, 10801 University Boulevard, Manassas, VA. The accession number indicated was assigned after successful viability testing, and the requisite fees were paid. made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of viable cultures for a period of thirty (30) years from the date of deposit. The organisms will be made available by the ATCC under the terms of the Budapest Treaty, which assures permanent and unrestricted availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 U.S.C. §122 and the Commissioner's rules pursuant thereto (including 37 C.F.R. §1.12 with particular reference to 886 OG 638). Upon the granting of a patent, all restrictions on the availability to the public of the deposited cultures will be irrevocably removed.

These deposits are provided merely as convenience to those of skill in the art, and are not an admission that a deposit is required under 35 U.S.C. §112. The nucleic acid sequences of these genes, as well as the amino acid sequences of the molecules encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the description herein. A license may be required to make, use, or sell the deposited materials, and no such license is hereby granted.

Plasmid	Deposit Date	ATCC No.
E1E2-809	August 16, 2001	PTA-3643

III. Experimental

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Those of skill in the art will readily appreciate that the invention may be practiced in a variety of ways given the teaching of this disclosure.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

EXAMPLE 1

Production of NS3NS4NS5a Polynucleotides.

A polynucleotide encoding NS3NS4NS5a (approximately amino acids 1027 to 2399, numbered relative to HCV-1) (also termed "NS345a" herein) or NS5a (approximately amino acids 1973 to 2399, numbered relative to HCV-1) was isolated from an HCV. Polynucleotides encoding a methionine residue were ligated to the 5' end of these polynucleotides and the polynucleotides were cloned into plasmid, vaccinia virus, and adenovirus vectors.

Immunization Protocols. In one immunization protocol, mice were immunized with 50 μg of plasmid DNA encoding either NS5a or encoding an NS3NS4NS5a fusion protein by intramuscular injection into the tibialis anterior. A booster injection of 10^7 pfu of vaccinia virus (VV)-NS5a (intraperitoneal) or 50 μg of plasmid control (intramuscular) was provided 6 weeks later.

In another immunization protocol, mice were injected intramuscularly in the tibialis anterior with 10^{10} adenovirus particles encoding an NS3NS4NS5a fusion protein. An intraperitoneal booster injection of 10^7 pfu of VV-NS5a or an intramuscular booster injection of 10^{10} adenovirus particles encoding NS3NS4NS5a was provided 6 weeks later.

EXAMPLE 2

Immunization with DNA encoding an NS3NS4NS5a fusion protein activates HCV-specific $CD8^+$ T cells.

^{51}Cr Release Assay. A ^{51}Cr release assay was used to measure the ability of HCV-specific T cells to lyse target cells displaying an NS5a epitope. Spleen cells were pooled from the immunized animals. These cells were restimulated *in vitro* for 6 days with the CTL epitopic peptide p214K9 (2152-HEYPVGSQL-2160; SEQ ID NO:1) from HCV-NS5a in the presence of IL-2. The spleen cells were then assayed for cytotoxic activity in a standard ^{51}Cr release assay against peptide-sensitized target cells (L929) expressing class I, but not class II MHC molecules, as described in Weiss (1980) J. Biol. Chem. 255:9912-9917. Ratios of effector (T cells) to target (B cells) of 60:1, 20:1, and 7:1 were tested. Percent specific lysis was calculated for each effector to target ratio.

The results of the assays are shown in Tables 1 and 2. Table 1 demonstrates that immunization with plasmid DNA encoding an NS3NS4NS5a fusion protein activates $CD8^+$ T cells which recognize and lyse target cells displaying an NS5a epitope. Surprisingly the NS5a polypeptide of the NS3NS4NS5a fusion protein was able to activate T cells even though the NS5a polypeptide was present in a fusion protein.

Similarly, Table 2 demonstrates that delivery of the NS3NS4NS5a fusion protein to mice by means of an adenovirus vector also activates $CD8^+$ T cells which recognize and lyse target cells displaying an HCV NS5a epitope. Thus, immunization with either "naked" (plasmid) DNA encoding an NS3NS4NS5a fusion protein or adenovirus vector-encoded fusion protein can be used to activate HCV-specific T cells.

EXAMPLE 3

Immunization with DNA encoding an NS3NS4NS5a fusion protein activates HCV-specific $CD8^+$ T cells which express IFN-γ.

Intracellular Staining for Interferon-gamma (IFN-γ). Intracellular staining for IFN-γ was used to identify the $CD8^+$ T cells that secrete IFN-γ after *in vitro* stimulation with the NS5a epitope p214K9. Spleen cells of individual immunized mice were restimulated *in vitro* either with p214K9 or with a non-specific peptide for 6-12 hours in the presence of IL-2 and monensin. The cells were then stained for surface CD8 and for intracellular IFN-γ and analyzed by flow cytometry. The percent of $CD8^+$ T cells which were also positive for IFN-γ was then calculated. The results of these assays are shown in Tables 1 and 2. Table 1 demonstrates that $CD8^+$ T cells activated in response to immunization with plasmid DNA encoding an NS3NS4NS5a fusion protein also express IFN-γ. Immunization with an NS3NS4NS5a fusion protein encoded in an adenovirus also results in $CD8^+$ HCV-specific T cells which express IFN-γ, although to a lesser extent than immunization with a plasmid-encoded NS3NS4NS5a fusion protein (Table 2).

Table 1. HCV-NS5a-Specific CD8+ T Cells in Mice Immunized with NS5a or NS345a DNA								
^{51}Cr Release Assay					Intracellular Staining for IFN-γ			
Percent Specific Lysis of Targets*					Percent of CD8+ T Cells Positive for IFN-g**			
E:T ratio	NS5a DNA		NS345a DNA		NS5a DNA		NS345a DNA	
	p214K9	-	p214K9	-	p214K9	-	p214K9	-
60:1	77	5	66	6				
20:1	61	4	49	2	1.74	0.26	1.18	0.40
7:1	29	1	29	4				

*Target cells (L929) were pulsed with p214K9 or media alone and labeled with ^{51}Cr.
**Spleen cells were cultured with p214K9 or media alone for 12 hours in the presence of monensin.

p214K9 is a CTL epitopic peptide (2152-HEYPVGSQL-2160, SEQ ID NO:1) from HCV-NS5a
'-' refers to the absence of peptide

Table 2. HCV-NS5a-Specific CD8+ T Cells Primed by Adenovirus or DNA Encoding for NS345a								
^{51}Cr Release Assay					Intracellular Staining for IFN-γ			
Percent Specific Lysis of Targets*					Percent of CD8+ T Cells Positive for IFN-g**			
E:T ratio	NS345a Adeno		NS345a DNA		NS345a Adeno		NS345a DNA	
	p214K9	-	p214K9	-	p214K9	p214J	p214K9	p214J
60:1	76	2	55	5				
20:1	85	2	22	3	3.24	0.13	0.25	0.09
7:1	62	<1	10	3				

*Target cells (L929) were pulsed with p214K9 or p214J and labeled with ^{51}Cr.
**Spleen cells were cultured with p214K9 or p214J for 12 hours in the presence of monensin.

p214K9 is a CTL epitopic peptide (2152-HEYPVGSQL-2160, SEQ ID NO:1) from HCV-NS5a
P214J is a control peptide (10 mer) from HCV-NS5a

EXAMPLE 4

Immunization with DNA encoding an NS3NS4NS5a fusion protein stimulates proliferation of HCV-specific $CD4^+$ T cells.

Lymphoproliferation assay. Spleen cells from pooled immunized mice were depleted of $CD8^+$ T cells using magnetic beads and were cultured in triplicate with either p222D, an NS5a-epitopic peptide from HCV-NS5a (2224-AELIEANLLWRQEMG-2238; SEQ ID NO:2), or in medium alone. After 72 hours, cells were pulsed with 1μ Ci per well of ^{3}H-thymidine and harvested 6-8 hours later. Incorporation of radioactivity was measured after harvesting. The mean cpm was calculated.

As shown in Table 3, immunization with a plasmid-encoded NS3NS4NS5a fusion protein stimulates proliferation of $CD4^+$ HCV-specific T cells. Immunization with an adenovirus vector encoding the fusion protein also resulted in stimulated proliferation of $CD4^+$ HCV-specific T cells (Table 4).

Table 3. HCV-NS5a-Specific CD4+ T Cells in Mice Immunized with NS5a or NS345a DNA

Mean CPM			
NS5a DNA		NS345a DNA	
p222D	media	p222D	media
4523	740	4562	861
(x6.1)		(x5.3)	

p222D is a CD4+ epitopic peptide (aa: 2224-AELIEANLLWRQEMG-2238, SEQ ID NO:2) from HCV-NS5a

Table 4. HCV-NS5-Specific CD4+ T Cells Primed by Adenovirus or DNA Encoding for NS345a			
Mean CPM			
NS345a Adeno		NS345a DNA	
p222D	media	p222D	media
896	357	1510	385
(x2.5)		(x3.9)	

p222D is a CD4+ epitopic peptide (aa: 2224-AELIEANLLWRQEMG-2238, SEQ ID NO:2) from HCV-NS5a

EXAMPLE 5

Efficiency of NS345a-encoding DNA Vaccine Formulations to prime CTLs in mice.

Mice were immunized with either 10-100 μg of plasmid DNA encoding NS345a fusion protein as described in Example 1, with PLG-linked DNA encoding NS345a, described below, or with DNA encoding NS345a, delivered via electroporation (see, e.g., U.S. Patent Nos. 6,132,419; 6,451,002, 6,418,341, 6233,483, U.S. Patent Publication No. 2002/0146831; and International Publication No. WO/0045823, for this delivery technique). The immunizations were followed by a booster injection 6 weeks later of 1 x 10^7 pfu vaccinia virus encoding NS5a, plasmid DNA encoding NS345a or plasmid DNA encoding NS5a each as described in Example 1.

PLG-delivered DNA. The polylactide-co-glycolide (PLG) polymers were obtained from Boehringer Ingelheim, U.S.A. The PLG polymer used in this study was RG505, which has a copolymer ratio of 50/50 and a molecular weight of 65 kDa (manufacturers data). Cationic microparticles with adsorbed DNA were prepared using a modified solvent evaporation process, essentially as described in Singh et al., *Proc. Natl. Acad. Sci. USA* (2000) 97:811-816. Briefly, the microparticles were prepared by emulsifying 10 ml of a 5% w/v polymer solution in methylene chloride with 1 ml of PBS at high speed using an IKA homogenizer. The primary emulsion was then added to 50ml of distilled water containing cetyl trimethyl ammonium

bromide (CTAB) (0.5% w/v). This resulted in the formation of a w/o/w emulsion which was stirred at 6000 rpm for 12 hours at room temperature, allowing the methylene chloride to evaporate. The resulting microparticles were washed twice in distilled water by centrifugation at 10,000 g and freeze dried. Following preparation, washing and collection, DNA was adsorbed onto the microparticles by incubating 100 mg of cationic microparticles in a 1mg/ml solution of DNA at 4 C for 6 hours. The microparticles were then separated by centrifugation, the pellet washed with TE buffer and the microparticles were freeze dried.

CTL activity and IFN-γ expression were measured by ^{51}Cr release assay or intracellular staining as described in examples 2 and 3 respectively. The results are shown in Table 5.

Results demonstrate that immunization using plasmid DNA encoding for NS345a to prime mice results in activation of CD8+ HCV specific T cells.

Table 5: Efficiency of NS345a-Encoding DNA Vaccine Formulations to Prime CTLs in Mice

			ICS for IFN-gamma (%CD8+ cells that are IFN-g+)					
NS345a DNA Vaccines	Boost	Mean	SdtdevP	# of mice tested	% respon d- ing	# of expts	fold increase vs. 'naked' DNA	CTL activ i- ty?
NS345a DNA	VVNS5a	1.02	1.70	41	68%	10	N/A	YES
NS345a DNA	NS345a DNA	0.02	0.04	22	5%	5	N/A	YES
NS345a DNA	NS5a DNA	0.22	0.21	24	63%	5	N/A	YES
NS345a DNA eV (electro-poration)	VVNS5a	5.00	4.36	7	100%	2	4.90	YES
PLGNS345a DNA	VVNS5a	2.65	2.54	6	100%	2	2.60	YES
PLGNS345a DNA	NS5a DNA	0.33	0.24	15	80%	3	1.50	YES

EXAMPLE 6

Immunization routes and replicon particles SINCR (DC+) encoding for NS345a

Alphavirus replicon particles, for example, SINCR (DC+) were prepared as described in Polo et al., *Proc. Natl. Acad. Sci. USA* (1999) 96:4598-4603. Mice were injected with 5 x 10^6 IU SINCR (DC+) replicon particles encoding for NS345a intramuscularly (IM) as described in Example 1, or subcutaneously (S/C) at the base of the tail (BoT) and foot pad (FP), or with a combination of 2/3 of the DNA delivered via IM administration and 1/3 via a BoT route. The immunizations were followed by a booster injection of vaccinia virus encoding NS5a as described in Example 1.

IFN-γ expression was measured by intracellular staining as described in Example 3. The results are shown in Table 6. The results demonstrate that immunization via SINCR (DC+) replicon particles encoding for NS345a by a variety of routes results in CD8+ HCV specific T cells which express IFN-γ.

Table 6: Immunization Routes and SINCR (DC+) Replicon Particles Encoding NS345a
(all mice VVNS5a challenged)

		ICS for IFN-gamma (%CD8+ cells that are IFN-g+)				
Vaccines	Immunization Route	Mean	SdtdevP	# of mice tested	# of expts	% responding mice
SINCR (DC+) $5X10^6$	100% IM (ta)	1.11	0.63	3	1	100%
SINCR (DC+) $5X10^6$	100% S/C (BoT + FP)	0.62	0.29	3	1	100%
SINCR (DC+) $5X10^6$	2/3 IM (ta) + 1/3 S/C (BoT)	2.43	2.00	3	1	100%

EXAMPLE 7

SINCR (DC+) vs SINDC (LP) replicon particles encoding for NS345a

Alphavirus replicon particles, for example, SINCR (DC+) and SINCR (LP) were prepared as described in Polo et al., *Proc. Natl. Acad. Sci. USA* (1999) 96:4598-4603. Mice were immunized with 1 x 10^3 to 1 x 10^7 IU of SINCR (DC+) or SINCR (LP) replicon particles encoding for NS345a, by intramuscular injection into the tibialis anterior, followed by a booster injection of 10^7 pfu vaccinia virus encoding NS5a at 6 weeks.

IFN-γ expression was measured by intracellular staining as described in Example 3. Administration of an increase in the number of SINCR (DC+) replicon particles encoding NS345a resulted in an increase in % of CD8+ T cells expressing IFN-γ.

EXAMPLE 8

Alphavirus replicon priming, followed by various boosting regimes.

Alphavirus replicon particles, for example, SINCR (DC+) were prepared as described in Polo et al., *Proc. Natl. Acad. Sci. USA* (1999) 96:4598-4603. Mice were primed with SINCR (DC+), 1.5 x 10^6 IU replicon particles encoding NS345a, by intramuscular injection into the tibialis anterior, followed by a booster of either 10-100 μg of plasmid DNA encoding for NS5a, 10^{10} adenovirus particles encoding NS345a, 1.5 x 10^6 IU SINCR (DC+) replicon particles encoding NS345a, or 10^7 pfu vaccinia virus encoding NS5a at 6 weeks.

IFN-γ expression was measured by intracellular staining as described in Example 3. The results are shown in Table 7. The results demonstrate that boosting with vaccinia virus encoding NS5a DNA results in the strongest generation of CD8+ HCV specific T cells which express IFN-γ. Boosting with plasmid encoding NS5a DNA also results in a good response, while lesser responses are noted with adenovirus NS345a or SINCR DC+ boosted animals.

Table 7: Alphavirus Replicon Particle Priming, Followed by Various Boosting Regimens						
		ICS for IFN-gamma (%CD8+ cells that are IFN-g+)				
Vaccines	Boost	Mean	SdtdevP	# of mice tested	# of expts	% respond-ing mice
SINCR (DC+) $1.5X10^6$	NS5a DNA	0.46	0.36	4	1	75%
SINCR (DC+) $1.5X10^6$	Adeno NS345a ($10X10^{10}$)	0.04	0.04	4	1	25%
SINCR (DC+) $1.5X10^6$	SINCR (DC+) $1.5X10^6$	0.06	0.06	8	2	25%
SINCR (DC+) $1.5X10^6$	VVNS5a ($1X10^7$)	2.43	2.45	4	1	100%

EXAMPLE 9

Alphaviruses expressing NS345a

Alphavirus replicon particles, for example, SINCR (DC+) and SINCR (LP) were prepared as described in Polo et al., *Proc. Natl. Acad. Sci. USA* (1999) 96:4598-4603. Mice were immunized with 1 x 10^2 to 1 x 10^6 IU SINCR (DC+) replicons encoding NS345a via a combination of delivery routes (2/3 IM and 1/3 S/C) as well as by S/C alone, or with 1 x 10^2 to 1 x 10^6 IU SINCR (LP) replicon particles encoding NS345a via a combination of delivery routes (2/3 IM and 1/3 S/C) as well as by S/C alone.

The immunizations were followed by a booster injection of 10^7 pfu vaccinia virus encoding NS5a at 6 weeks.

IFN-γ expression was measured by intracellular staining as described in Example 3. The results are shown in Figure 6. The results indicate activation of CD8+ HCV specific T cells.

EXAMPLE 10

Efficiency of NS5a encoding DNA vaccine formulations to prime CTLs in mice

Mice were immunized with either 10-100 μg of plasmid DNA encoding NS5a as described in Example 1 or with PLG-linked DNA encoding NS5a as described in Example 5. The immunizations were followed by a booster injection at 6 weeks of either 10-100 μg of plasmid DNA encoding for NS5a, 10^{10} adenovirus particles encoding NS345a, 1.5 x 10^6 IU SINCR (DC+) replicon particles encoding NS345a, or 10^7 pfu vaccinia virus encoding NS5a.

CTL activity and IFN-γ expression were measured by the methods described in Examples 2 and 3.

The results are shown in Table 8. The results demonstrate that priming with plasmid DNA encoding for NS5a or PLG-linked DNA encoding NS5a results in activation of CD8+ HCV specific T cells.

Table 8: Efficiency of NS5a-Encoding DNA Vaccine Formulations to Prime CTLs in Mice

			ICS for IFN-gamma (%CD8+ cells that are IFN-g+)						
NS5a Vaccines	Boost	Mean	SdtdevP	# of mice tested	% respond-ing	# of expts	fold increase vs. 'naked' DNA	CTL activity ?
NS5a DNA	VVNS5a	1.67	1.49	8	100%	3	N/A	YES
NS5a DNA	NS5a DNA	0.17	0.09	12	83%	3	N/A	YES
PLGNS5a DNA	NS5a DNA	0.22	0.09	9	100%	2	1.29	YES
NS5a DNA	AdenoNS 345a	0.10	0.08	4	50%	1	N/A	NO
NS5a DNA	SINCRNS 345a	0.20	0.17	4	75%	1	N/A	YES

EXAMPLE 11

Efficiency of NS345b-encoding DNA vaccine formulations to prime CTLs in mice

Mice were immunized with 10-100 μg of plasmid DNA encoding NS34b by intramuscular injection to the tibialis anterior or with PLG linked DNA encoding NS5a as described in Example 5. The immunizations were followed by a booster injection of plasmid DNA encoding for NS5a as described in Example 1.

CTL activity and IFN-γ expression were measured by the methods described in Examples 2 and 3.

The results are shown in Table 9. The results demonstrate that priming with plasmid DNA encoding NS345b or PLG-linked NS345b results in activation of CD8+ HCV specific T cells.

Table 9: Efficiency of NS345b-Encoding DNA Vaccine Formulations to Prime CTLs in Mice

			ICS for IFN-gamma (%CD8+ cells that are IFN-g+)					
NS345 DNA Vaccines	Boost	Mean	SdtdevP	# of mice tested	% responding	# of expts	fold increase vs. 'naked' DNA	CTL activity?
NS345 DNA	NS5a DNA	0.18	0.16	15	60%	3	N/A	YES
PLGNS345 DNA	NS5a DNA	0.30	0.33	14	57%	3	1.67	YES

EXAMPLE 12

Administration of DNA via separate plasmids

Mice were immunized with 100 μg plasmid DNA encoding for NS345a or with 100 μg PLG-linked DNA encoding NS345a. Additionally, separate DNA plasmids encoding NS5a, NS34a, and NS4ab (33.3 μg each) were administered concurrently to another group of mice. Finally, PLG-linked DNA encoding NS5a, NS34a, and NS4ab (33.3 μg each) were administered concurrently to another group of mice. The immunizations were followed by a booster injection of $1x10^7$ pfu vaccinia virus encoding NS5a, 6 weeks post first immunization.

IFN-γ expression was measured by the method described in Example 3. The results are shown in Figure 7. The results demonstrate a particularly vigorous response in the activation of CD8+ HCV specific T cells when the DNA is broken down into smaller sub units and linked to PLG.

EXAMPLE 13

Immunogenicity of NS345Core$_{121}$-ISCOMS in Mice

Groups of 10 C57 black mice were immunized IM at 0, 21 and 60 days with the formulations shown in Table 10. The NS345Core$_{121}$-PLGdss group received a vaccine dose of 50 μl in each leg whereas the other vaccine groups received a vaccine dose of 50 μl in one leg.

NS345Core$_{121}$-ISCOMS were comprised of amino acids 1242 to 3011 and 1-121 and the HCV polyprotein, numbered relative to HCV-1and were adsorbed to ISCOMS with a ratio of protein to QH of approximately 8:1, using standard techniques. See, e.g., International Publication No. WO 01/37869A.

Core-ISCOMS including an HCV core protein fragment from the region spanning amino acid positions 1-191 of the HCV polyprotein, numbered relative to HCV-1, with a ratio of protein to QH of 1:1, were produced using standard techniques. See, e.g., International Publication No. WO 01/37869A.

NS345Core$_{121}$ was formulated in 0.1% SDS in PBS and contained DTT. Protein was diluted in PBS and mixed 1:1 with MF59 (see, Ott et al., "MF59 –

Design and Evaluation of a Safe and Potent Adjuvant for Human Vaccines" in *Vaccine Design: The Subunit and Adjuvant Approach* (Powell, M.F. and Newman, M.J. eds.) Plenum Press, New York (1995) pp. 277-296; and U.S. Patent No. 6,299,884) prior to immunization.

For NS345Core$_{121}$-PLGdss, PLG microparticles produced as described above were treated with 3-(trimethylsilyl)-1-propanesulfonic acid (DSS) to enhance adsorption of antigen. DSS is commercially available from, e.g., Sigma Chemical Co., St. Louis, MO. NS345Core$_{121}$ was adsorbed thereto using standard techniques (see, International Publication No. WO 00/050006). The NS345Core$_{121}$-PLGdss was mixed with MF59 prior to immunization.

As shown in Table 10, NS345Core$_{121}$-ISCOMS produced antibody response only to NS5 in immunized C57 black mice. Higher levels of antibodies to NS5 were produced in mice immunized with NS345Core$_{121}$ adjuvanted with MF59, however no antibody response to core, NS3 or NS4 was produced with this adjuvant either.

Mice immunized with Core-ISCOMS produced antibodies to core. In contrast, NS345Core$_{121}$-PLGdss immunized mice produced significantly higher antibodies to NS5 than NS345Core$_{121}$-ISCOMS. In addition, NS345Core$_{121}$-PLGdss immunized mice produced antibodies to NS3 and some antibody response to core, but no antibodies to NS4.

Table 10. Immunogenicity of NS345Core$_{121}$-Iscoms in Mice. Geometric mean EIA antibody titers to core and nonstructural proteins are shown. The number of responding mice per group are also listed.

Vaccine	IM Protein Dose (μg)[a]	Anti-Core Antibody EIA GMT	Anti-C33C (NS3) Antibody EIA GMT	Anti-C100 (NS4) Antibody EIA GMT	Anti-NS5 Antibody EIA GMT
NS345Core$_{121}$ ISCOMS	6.0, 6.0, 6.0[b]	<10 (0/10)	<10 (0/10)	<10 (0/10)	31 (7/10)
Core-ISCOMS	6.0, 6.0, 6.0[c]	188 (9/10)	<10 (1/10)	<10 (0/10)	<10 (2/10)
NS345Core$_{121}$ MF59	6.0, 6.0, 6.0	<10 (2/9)	<10 (1/9)	<10 (0/9)	279 (9/)
NS345Core$_{121}$ PLG-dss/MF59	10, 10, 10	5 (6/10)	50 (9/10)	<10 (2/10)	419 (9/9)

[a]Groups of 10 C57 black mice were immunized IM at 0, 21 and 60 days. Serum was obtained after the last immunization. The NS345 Core$_{121}$-PLGdss group received vaccine dose of 50 μl in each leg whereas the other vaccine groups received vaccine dose of 50 μl in one leg.
[b]The ratio of protein to QH was approximately 8:1.
[c]The ratio of protein to QH was approximately 1:1.

EXAMPLE 14

Immunogenicity of Different Formulations of NS345Core$_{121}$ or NS345 in Mice

Groups of 10 C57 black mice were immunized IM at 0, 30 and 60 days with the formulations shown in Tables 11 and 12. For the studies shown in Table 11, the NS345 and NS345Core$_{121}$ protein concentration was 10 μg per dose, and for those in Table 12, the concentration was 5 μg per dose.

For PLG-NS345 (amino acids 1242 to 3011 of the HCV polyprotein) and PLG-NS345Core$_{121}$ (amino acids 1242-3011 and 1-121 of the HCV polyprotein), PLG microparticles were prepared and NS345 or NS345Core$_{121}$ were adsorbed thereto using standard techniques, as described above.

For PLG-NS345 + PLG-CTAB-E1E2 DNA, PLG microparticles were prepared and NS345 was adsorbed to the microparticles as described above. E1E2 DNA was produced as follows. Mammalian expression plasmid pMH-E1E2-809 (Figure 4, ATCC Deposit No. PTA-3643) encodes an E1E2 fusion protein which includes amino acids 192-809 of HCV 1a (see, Choo et al., *Proc. Natl. Acad. Sci. USA* (1991) 88:2451-2455). Chinese Hamster Ovary (CHO) cells were used for expression of the HCV E1E2 sequence from pMH-E1E2-809. In particular, CHO DG44 cells were used. These cells, described by Uraub et al., *Proc. Natl. Acad. Sci. USA* (1980) 77:4216-4220, were derived from CHO K-1 cells and were made dihydrofolate reductase (dhfr) deficient by virtue of a double deletion in the dhfr gene. DG44 cells were transfected with pMH-E1E2-809. The transfected cells were grown in selective medium such that only those cells expressing the dhfr gene could grow (Sambrook et al., supra). Isolated CHO colonies were picked (~800 colonies) into individual wells of a 96-well plate. From the original 96-well plates, replicates were made to perform expression experiments. The replicate plates were grown until the cells made a confluent monolayer. The cells were fixed to the wells of the plate and permeablized using cold methanol. Anti-E1 and anti-E2 antibodies, 3D5C3 and 3E5-1 respectively, were used to probe the fixed cells. After adding an anti-mouse HRP conjugate, followed by substrate, the cell lines with the highest expression were determined. The highest expressing cell lines were then expanded to 24-well cluster plates. The assay for expression was repeated, and again, the highest expressing cell lines were expanded to wells of greater volume. This was repeated until the highest expressing cell lines were expanded from 6-well plates into tissue culture flasks. At this point there was sufficient quantity of cells to allow accurate count and harvest of the cells, and quantitative expression assays were done. An ELISA was performed on the cell extract, to determine high expressors.

To produce the PLG-CTAB-E1E2 DNA, PLG microparticles were treated with CTAB as described above (see, International Publication No. WO 00/050006).

For PLG-NS345Core$_{121}$ + E1E2 DNA PLG-NS345Core$_{121}$ and E1E2 DNA were produced as described above.

For PLG-NS345 or PLG-NS345Core$_{121}$ + MF59, PLG-NS345 or PLG-NS345Core$_{121}$ was combined with MF59 as described above.

For PLG-NS345 or PLG-NS345Core$_{121}$ + CTAB-CpG, NS345 or NS345Core$_{121}$ was adsorbed to PLG as described above. The CpG molecule used was 5'-TCCATGACGTTCCTGACGTT-3' and this was treated with CTAB, as described above.

For PLG-NS345 or PLG-NS345Core$_{121}$ + QS21, the saponin adjuvant QS21 was combined with the PLG-HCV proteins.

For PLG-NS345 or PLG-NS345Core$_{121}$ + CTAB-CpG + MF59, the various components, as described above, were combined.

The remaining adjuvants used in the studies and shown in the tables are self-explanatory.

The results of these studies are shown in Tables 11 and 12. As can be seen in Table 11, none of the formulations produced antibody responses to core, NS3 or NS4 antigens. However , PLG-NS345+CTAB-CPG in MF59 produced the highest antibody titers to NS5. PLG-NS345Core$_{121}$+QS21, PLG-NS345+CTAB-CPG, PLG-NS345Core$_{121}$+CTAB-CPG, and PLG-NS345 + QS21 produced moderate antibody titers to NS5. The other formulations produced very low antibody titers to NS5.

As can be seen in Table 12, NS345Core$_{121}$ /MF59/MPL and NS345Core$_{121}$/MF59/CpG formulations produced very high antibody titers to NS345Core$_{121}$. NS345Core$_{121}$/MF59, NS345/MF59/CpG, and NS345Core$_{121}$ /MF59/Chol/QS21 formulations produced moderate antibody titers to NS345Core$_{121}$. The other formulations produced very low or no antibody titers to NS345Core$_{121}$.

Table 11. Immunogenicity of different formulations of HCV NS345Core$_{121}$ or NS345 in Mice. Geometric mean EIA antibody titers to core and nonstructural proteins are shown.

Vaccine[a]	Anti-Core Antibody EIA GMT	Anti-C33C (NS3) Antibody EIA GMT	Anti-C100 (NS4) Antibody EIA GMT	Anti-NS5 Antibody EIA GMT
PLG-NS345	<10	<10	<10	10
PLG-NS345Core$_{121}$	<10	<10	<10	15
PLG-NS345+PLG-CTAB-E1E2 DNA	<10	11	<10	23
PLG-NS345Core$_{121}$ +E1E2 DNA	<10	<10	<10	20
PLG-NS345 +MF59	<10	<10	<10	70
PLG-NS345Core$_{121}$ + MF59	<10	<10	<10	26
PLG-NS345 + CTAB-CPG	<10	<10	<10	350
PLG-NS345Core$_{121}$ + CTAB-CPG	<10	<10	<10	271
PLG-NS345 +QS21	<10	<10	<10	201
PLG-NS345Core$_{121}$ + QS21	<10	<10	<10	505

PLG-NS345 + CTAB-CPG + MF59	<10	<10	<10	1471
PLG-NS345Core$_{121}$ + CTAB+MF59	<10	<10	<10	63

[a]Groups of 10 C57 black mice were immunized IM at 0, 30 and 60 days. Serum was obtained after the last immunization. The NS345 or NS345Core$_{121}$ protein concentration was 10 μg per dose.

Table 12. Immunogenicity of different formulations of HCV NS345Core$_{121}$ or HCV NS345 in Mice. Geometric mean EIA antibody titers to NS345Core$_{121}$ protein are shown.

Vaccine[a]	Anti-NS345Core$_{121}$ Antibody EIA GMT
NS345Core$_{121}$/MF59	328
NS345Core$_{121}$/MF59/CpG	7,926
NS34a+NS5B+Core/MF59	12
NS34a+NS5B+Core/MF59/CpG	5
PLG-NS345Core$_{121}$/MF59	<10
PLG-NS345Core$_{121}$/MF59/CpG	<10
PLG-NS345Core$_{121}$/PLG-CpG	9
NS345Core$_{121}$/alum phosphate	34
NS345Core$_{121}$/alum phosphate//CpG	950
NS345/MF59/CpG	511
PLG-NS345/PLG/CpG	117
NS345Core $_{121}$/MF59/MPL	10,292
NS345Core $_{121}$/MF59/Chol/QS21	698
NS345Core $_{121}$/Alum phosphate/MPL	23

[a]Groups of 10 C57 black mice were immunized IM at 0, 30 and 60 days. Serum was obtained after the last immunization. The NS345 or NS345Core$_{121}$protein concentration was 5 μg per dose.

EXAMPLE 15

Lymphoproliferative Response of Different Formulations of NS345Core$_{121}$ or NS34A + NS35B + Core in Mice

Groups of 8 C57 black mice were immunized IM at 0, 30 and 60 days with the formulations shown in Table 13 and are as described above. Spleens were obtained after the last immunization. The NS345Core$_{121}$ protein concentration was 25 μg per dose. The NS34a, NS5b and core doses were 3 μg each.

The results of this study are shown in Table 13. As can be seen, NS345Core$_{121}$/Alum/CpG, PLG-NS345Core$_{121}$/PLG/CpG, NS34a+NS5B+ Core/MF59/CpG and PLG-NS345Core$_{121}$/MF59/CpG formulations demonstrated strong LPA responses to NS5, NS34 and core antigens. The NS345Core$_{121}$/MF59 formulation also produced a strong LPA response to NS5 and NS34. Core was not tested.

Moderate LPA responses were observed to NS5, NS34 and Core antigens with PLG-NS345Core$_{121}$/MF59 and NS34a +NS5B + Core/MF59 formulations. The NS345Core$_{121}$/MF59/CpG formulation may not have been administered properly in that no LPA response was observed in this experiment. In a subsequent experiment as shown in Table 14, an LPA was observed to this formulation.

Groups of 8 C57 black mice were immunized once IM with the formulations shown in Table 14, produced as described above. Draining lymph nodes were obtained.

The NS345Core$_{121}$ protein concentration was 25 μg per dose.

The results of this study are shown in Table 14. As can be seen in Table 14, all the formulations tested produced a strong LPA response to NS5, NS34 and Core as well as the NS345Core$_{121}$.

Table 13. Lymphoproliferative response of different formulations of HCV NS345Core$_{121}$ or NS34A+NS5B+Core in Mice. LPA responses (cpm) to core and nonstructural proteins are shown. The number of mice in each group responding is also indicated in parentheses.

Vaccine[a]	SOD-NS5	SOD-C200 (NS34)	SOD-C22-3 (Core)	HIV-2 env (background control)
NS345Core$_{121}$/MF59	2250 (6/8)	1800 (4/8)	ND	144
NS345Core$_{121}$/MF59/CpG	80	80	ND	138
PLG-NS345Core$_{121}$/MF59	560 (2/8)	120 (2/8)	510 (2/8)	93
PLG-NS345Core$_{121}$/MF59/CpG	1600 (6/8)	1500 (6/8)	620 (8/8)	75
NS34a+NS5B+Core/MF59	564 (8/8)	710 (8/8)	265 (8/8)	76
NS34a+NS5B+ Core/MF59/CpG	1523 (8/8)	885 (8/8)	446 (6/8)	67
PLG-NS345Core$_{121}$/PLG/CpG	3675 (8/8)	2860 (8/8)	370 (8/8)	88
NS345Core$_{121}$/Alum/CpG	8450 (8/8)	7940 (8/8)	1040 (6/8)	82

[a]Groups of 8 C57 black mice were immunized IM at 0, 30 and 60 days. Spleens were obtained after the last immunization. The NS345Core$_{121}$ protein concentration was 25 μg per dose. The NS34a, NS5B and Core doses were 3 μg each.

Table 14. Lymphoproliferative response of different formulations of HCV NS345Core$_{121}$ in Mice. The LPA responses (cpm) form an average of three consecutive experiments to core and nonstructural proteins are shown.

Vaccine[a]	SOD-NS5	SOD-C200 (NS34)	SOD-C22-3 (Core)	NS345Core$_{121}$	HIV-2 env (background control)
NS345Core$_{121}$/MF59	8900	8300	3000	20900	890
NS345Core$_{121}$/MF59/CpG	1890	1628	1200	20100	623
PLG-NS345Core$_{121}$/MF59	10700	12900	1800	23700	818
PLG-NS345Core$_{121}$/MF59/CpG	3600	4690	1660	27300	911
PLG-NS345Core$_{121}$	4500	4660	760	18150	315
PLG-NS345Core$_{121}$/PLG/CpG	7750	5300	1250	22980	450
NS345Core$_{121}$/Alum	3050	3725	744	11300	390
NS345Core$_{121}$/Alum/CpG	4130	4670	600	20660	480

[a]Groups of 8 C57 black mice were immunized once IM. Draining lymph nodes were obtained. The NS345Core$_{121}$ protein concentration was 25 µg per dose.

EXAMPLE 16

Immunogenicity of Recombinant HCV Protein Vaccines Adjuvanted with ISCOMS in Rhesus Macaques

The safety and immunogenicity of HCV proteins completed with the adjuvant, Iscomatrix, was studied in Rhesus macaques. Three groups made up of four animals each were immunized IM as detailed below at week 0, 4 and 8 weeks. Vaccines were prepared as described above. The ISCOMS used lacked QH-A.

Group Number	n	Vaccine	Delivery
1	4	Core-ISCOM (50 μg in 1 ml)	0.5 ml R Leg 0.5 ml L Leg
2	4	NS345Core$_{121}$-ISCOM (1 mg in 1 ml)	0.5 ml R Leg 0.5 ml L Leg
3	4	Core-ISCOM (25 μg in 0.5 ml) and NS5b-ISCOM (50 μg in 0.35 ml)	0.5 ml Core-ISCOM R Leg 0.35 ml NS5b-ISCOM L Leg

Bleeds occurred as follows and immunogenicity was determined by CTL assays, lymphoproliferation assays, FACS analysis and antibody response a previously described (Palakos, et al. (2001) *J. of Immunology* 166:3589).

Week	Bleed date	Immunized
-10		
-1	X	
0		X
2	X	
4		X
6	X	
8		X
10	X	

The immunogenicity of the different HCV recombinant protein vaccines is shown in Tables 15-17.

Table 15. The Immunogenicity of HCV Core-ISCOMS vaccine two weeks post 2nd immunization and post 3rd immunization as assessed by CTL assays, CD8+ FACS analysis, LPA stimulation index and CD4+ FACS analysis

	2 weeks post 3°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X020	- (-)					+(-)				
N001	- (-)					+(-)				
N086	- (-)					+(-)				
X010	- (-)					+(11)				

	2 weeks post 2°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X020	- (-)					+/-(-)				
N001	- (-)					-(8)				
N086	- (-)					+(-)				
X010	- (-)					+/- (12)				

Table 16. The Immunogenicity of HCV NS345Core$_{121}$-ISCOMS vaccine two weeks post 2nd immunization and post 3rd immunization as assessed by CTL assays, CD8+ FACS analysis, LPA stimulation index and CD4+ FACS analysis

	2 weeks post 3°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X016	- (-)	+(+)	-(-)	-(-)	-(-)	-(-)	+(-)	-(-)	+/-(-)	+/-(-)
X008	- (-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(5)	-(-)
X021	- (-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)	-(-)
X023	+/- (-)	-(-)	-(-)	+/-(-)	-(-)	-(-)	+/-(-)	-(-)	-(-)	-(-)

	2 weeks post 2°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X016	- (-)	+(+)	-(-)	+(+)	+(+)	-(-)	+(-)	+/-(-)	+(-)	+(-)
X008	+ (-)	+(+)	-(-)	+(+)	+(+)	-(-)	+(-)	-(-)	+-(-)	+(-)
X021	- (-)	-(-)	+/-(-)	-(-)	+/-(-)	-(-)	-(-)	+/-(-)	-(-)	-(-)
X023	+/- (+)	+(+)	-(-)	+/-(-)	+(-)	-(-)	+(-)	-(-)	-(-)	7

Table 17. The Immunogenicity of HCV Core-ISCOMS + NS5b-ISCOMS vaccine two weeks post 2nd immunization and post 3rd immunization as assessed by CTL assays, CD8+ FACS analysis, LPA stimulation index and CD4+ FACS analysis

	2 weeks post 3°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X022	+(-)				-(-)	-(8)				+/-(11)
X014	-(-)				-(-)	+(6)				+(11)
N154	-(-)				-(-)	-(-)				+(-)
N173	-(-)				-(-)	-(-)				+/-(-)

	2 weeks post 2°									
Macaque#	CD8+ ICS (CTL)					CD4+ ICS (LPA SI)				
	C	NS3	NS4	NS5a	NS5b	C	NS3	NS4	NS5a	NS5b
X022	-(-)				-(+)	-(-)				-(-)
X014	+(-)				+/-(-)	-(-)				+(-)
N154	-(-)				-(-)	-(6)				+(8)
N173	-(-)				-(-)	+/-(-)				+(6)

As can be seen in Table 15, the HCV Core-ISCOM vaccine produced no CTL positive responses in any of the 4 immunized macaques after the second or third immunizations. No positive CD8 γ-interferon and/or TNF-α intracellular staining was also observed, although backgrounds were high in these particular arrays. At least two of four macaques produced a strong LPA response after the second immunizations, but only one remained positive after the third immunization. Two of four macaques produced positive CD4 intracellular staining after the second immunization and four of four after the third immunization.

As shown in Table 16, the HCV NS345Core$_{121}$-ISCOM vaccine after the second immunization produced CTL positive responses to peptide pools representing two or more HCV proteins in three of four macaques (two of these macaques had responses to peptide pools from NS3, NS5a and NS5b, one to peptide pools from core and NS3). CD8 positive γ-interferon and/or TNF-α intracellular staining to peptide pools representing two or more HCV proteins was positive in at least three of four macaques. One of four macaques produced a strong LPA response. At least three of four macaques produced CD4 positive intracellular staining to two or more HCV proteins. After the third immunization, only one of four macaques had a positive CTL response, CD8 positive intracellular staining and C04 positive intracellular staining. One other macaque had a positive LPA response and weak CD8+ CD4 intracellular staining, This decline in immunogenicity was likely due to instability of the vaccine formulation (see below).

As shown in Table 17, the HCV Core-ISCOM + NS5b-ISCOM vaccine produced a CTL positive response to NS5b in one of the 4 immunized macaques after the second immunization which did not remain positive after the third immunizaton. CD8 positive intracellular positive staining was observed in one of four animals post second. Two of four macaques produced a strong LPA response after the second immunization which did not remain positive after the third immunization. Two other macaques did develop a strong LPA response after the third immunization. Three or four developed positive CD4 intracellular staining. One developed positive CD8 intracellular staining.

Three weeks after the third immunization, it was noted that the physical appearance of the polyprotein vaccine solution was visibly turbid. The core vaccine also was turbid but less so. The Core-NS5 vaccine was also slightly turbid. Analysis of this turbidity in the polyprotein formulation indicated that the ISCOM particles had precipitated into large aggregates. These aggregates could be dispersed by vortexing with 0.1% TWEEN 80 detergent. It is probable that this change in the formulation of the vaccine occurred before the last immunization. This observed change in appearance of the vaccines may have affected their immunogencity as cellular immune results declined in all three vaccines.

The immunogenicity of HCV Core-ISCOMS, NS345Core$_{121}$-ISCOMS and Core-ISCOMS + NS5b-ISCOMS as assessed by EIA antibody response is shown in Table 18. As can be seen, all three vaccines produced an antibody response by the third immunization to their corresponding HCV proteins, except for the NS345Core$_{121}$-ISCOM vaccine. The NS345Core$_{121}$-ISCOM vaccine produced antibody responses to NS3, NS4 and a very strong antibody response to NS5, but no antibody response to HCV core.

Table 18. The immunogenicity of HCV Core-ISCOMS, NS345Core$_{121}$-ISCOMS, Core-ISOCMS + NS5b-ISCOMS vaccine two weeks post 2nd immunization and post 3rd immunization as assessed by EIA antibody response to HCV proteins.

Vaccine Macaque #	Anti-Core EIA Antibody Titer		Anti-NS3 EIA Antibody Titer		Anti-NS4 EIA Antibody Titer		Anti-NS5 EIA Antibody Titer	
	Post 2nd	Post 3rd	Post 2nd	Post 3rd	Post 2nd	Post 3rd	Post 2nd	Post 3rd
Core-ISCOM								
X020	66	226						
N001	87	46						
N086	363	396						
X010	108	137						
NS345 Core121/ ISCOM								
X016	<10	<10	<10	554	56	68	3,590	3,405
X008	<10	<10	66	995	14	44	2,109	3,213
X021	<10	<10	128	6,330	41	204	7,213	8,083
X023	<10	<10	<10	3,910	64	64	1,243	4,704
Core-ISCOM + NS5b-ISCOM								
X022	<10	18					<10	134
X014	<10	13					<10	693
N154	542	554					<10	272
N173	28	78					<10	258

EXAMPLE 17

Immunization of Chimpanzees with Recombinant HCV Protein and DNA Vaccines

Five groups of five chimps each were immunized IM at 0, 0.7, 2 and 5 months with the formulations presented below. Blood was collected at week 0, two weeks subsequent to the second immunization, two weeks following the third immunization and two weeks after the fourth immunization.

Formulation 1: 20 μg E1E2 polypeptide + MF59 + 500 μg CpG (produced as described above);

Formulation 2: 1 mg NS345Core$_{121}$-ISCOM (produced as described above);

Formulation 3: 6 mg each of CTAB-PLG-E1E2 (bp 574-2427, encoding amino acids 192-809 of the HCV polyprotein, numbered relative to HCV-1); CTAB-PLG-NS34a (bp 3079-5133, encoding amino acids 1027-1711 of the HCV polyprotein, numbered relative to HCV-1); CTAB-PLG-NS34ab (bp 4972-5916, encoding amino acids 1658-1972 of the HCV polyprotein, numbered relative to HCV-1); CTAB-PLG-NS5a (bp 5917-7260, encoding amino acids 1973-2420 of the HCV polyprotein, numbered relative to HCV-1);

Formulation 4: 6 mg each of E1E2 DNA, NS34a DNA, NS34ab DNA and NS5a DNA, having the same coordinates as described above, delivered without PLG via electroporation (see, e.g., U.S. Patent Nos. 6,132,419; 6,451,002, 6,418,341, 6233,483, U.S. Patent Publication No. 2002/0146831; and International Publication No. WO/0045823, for this delivery technique). Results are shown in Figures 8-10.

As can be seen, in Figure 8, all vaccines were capable of priming CD4+ and CD8+ cells specific to HCV. Thus, all vaccines were successful at inducing a T cell response to HCV. Determination of the results for the PLG-DNA from formulation 3 at two weeks subsequent to the fourth vaccination is in progress.

As shown in Figures 9 and 10, multiple T cell specificities were induced by the two vaccines. Both vaccines primed T-cells specific for multiple T cell epitopes.

As can be seen in Tables 19 and 20, E1E2 adjuvanted with MF59 primed

anti-E1E2 titers. CpG enhanced anti-E1E2 responses as well as TH1 responses and the ISCOM and the two DNA vaccines were capable of priming CD4+ and CD8+ T cell responses to HCV.

Table 19. Anti-E1E2 EIA antibody titers in chimps immunized with Electroporated DNA E1E2NS345a or PLG DNA E1E2NS345					
Vaccine	Chimp	Pre 1st	Post 2nd	Post 3rd	Post 4th
Electroporated DNA	4X0330	-	-	-	9
E1E2-NS34A-	4X0335	-	-	-	10
NS4AB-NS5A[a]	4X0348[c]	10	457	198	50
	4X0354[d]	206	1,261	1,197	207
	4X0368[c]	245	1,426	1,267	358
PLG DNA	4X0238	-	-	30	10
E1E2-NS34A-	4X0239	-	104	309	-
NS4AB-NS5A[b]	4X0250	-	-	12	-
	4X0278	-	-	29	-
	4X0288	-	-	12	-

[a]Electroporated IM with 1.5 mg of each plasmid at 0, 0.7, 2 and 5 months.
Bleeds were taken 14 days after each immunization.
[b]IM immunization with 1.5 mg of each PLG plasmid at 0, 0.7, 2 and 6 months.
Bleeds were taken 14 days after each immunization.
[c]Prior E2 immunization
[d]Prior E1E2 immunization

Table 20. Immunogenicity in chimps of low dose (20 μg) HCV E1E2 antigen using MF59 or MF59 combined with CpG as adjuvants (2 wks post 3rd)				
Vaccine[a]	Chimp	E1E2EIA Ab Titer	E1E2 EIA Ab GMT	CD4+ (ICS)
E1E2/ MF59	4x0419	84		-
	4x0420	101		-
	4x0431	131	261	-
	4x0371	421		-
	4x0372	2,580		+/-
			P = 0.029[b]	
E1E2/ MF59/CpG	4x0410	8,835		-
	4x0426	2,713		-
	4x0365	3,201	2,713	+
	4x0367	510		-
	4x0346	1,238		++

[a]Chimps immunized IM at 0, 1 and 6 mos with 20 μg of E1E2 antigen using MF59 with or without 500 μg of CpG. Serum samples were obtained 14 days after last immunization.
[b]Chimps immunized with E1E2 using CpG combined with MF59 as adjuvant produced significantly higher ($P<0.05$) levels of E1E2 EIA antibody than chimpanzees with E1E2 using MF59 alone.*

Thus, HCV polypeptides and polynucleotides, either alone or as fusions, to stimulate cell-mediated immune responses, are disclosed. Although preferred embodiments of the subject invention have been described in some detail, it is understood that obvious variations can be made without departing from the spirit and the scope of the invention as defined by the appended claims.

We claim:

1. A fusion protein comprising HCV polypeptides, wherein the HCV polypeptides consist essentially of an NS3, an NS4, an NS5 and a core polypeptide of a hepatitis C virus (HCV), wherein said core polypeptide consists of amino acids 1-121 of the HCV polyprotein, numbered relative to the full-length HCV-1 polyprotein.

2. The fusion protein of claim 1, wherein the NS5 polypeptide is an NS5a polypeptide.

3. The fusion protein of claim 1, wherein the NS5 polypeptide is an NS5b polypeptide.

4. The fusion protein of claim 1, wherein the NS5 polypeptide is an NS5a and an NS5b polypeptide.

6. The fusion protein of claim 4, wherein the protein comprises the sequence of amino acids of SEQ ID NO:8.

7. A fusion protein according to any of claims 1-6, wherein at least one of the HCV polypeptides is derived from a different strain of HCV than the other HCV polypeptides.

8. A composition comprising:
(a) a fusion protein according to any of claims 1-7; and
(b) a pharmaceutically acceptable excipient.

9. The composition of claim 8, further comprising an adjuvant.

10. The composition of claim 8, further comprising a CpG oligonucleotide.

11. The composition of claim 8, wherein said fusion protein is adsorbed to or entrapped within a microparticle or ISCOM.

12. The composition of claim 8, further comprising a polynucleotide encoding an HCV E1E2 complex.

13. An isolated and purified polynucleotide that encodes a fusion protein according to any of claims 1-7.

14. A composition comprising:
(a) the isolated and purified polynucleotide of claim 13; and
(b) a pharmaceutically acceptable excipient.

15. The composition of claim 14, further comprising an adjuvant.

16. The composition of claim 14, wherein said polynucleotide is adsorbed to or entrapped within a microparticle.

17. The composition of claim 14, further comprising a polynucleotide encoding an HCV E1E2 complex.

18. A method of activating T cells of a vertebrate subject which recognize an epitope of an HCV polypeptide, comprising the step of:
administering the composition of any of claims 8-12 to said vertebrate

subject, whereby a population of activated T cells recognizes an epitope of the NS3, NS4, NS5 and/or core polypeptides.

19. A method of activating T cells of a vertebrate subject which recognize an epitope of an HCV polypeptide, comprising the step of:

administering the composition of any of claims 14-17 to said vertebrate subject, whereby a population of activated T cells recognizes an epitope of the NS3, NS4, NS5 and/or core polypeptides.

20. The method of claim 19 , wherein the polynucleotide is administered via electroporation.

21. Use of a composition according to any of claims 8-12 and 14-17 for activating T cells of a vertebrate subject which recognize an epitope of an HCV polypeptide, wherein a population of activated T cells recognizes an epitope of the NS3, NS4, NS5 and/or core polypeptides.

22. Use of a fusion protein according to any of claims 1-7 for the manufacture of a medicament for activating T cells of a vertebrate subject which recognize an epitope of an HCV polypeptide, wherein a population of activated T cells recognizes an epitope of the NS3, NS4, NS5 and/or core polypeptides.

23. Use of a polynucleotide according to claim 13 for the manufacture of a medicament for activating T cells of a vertebrate subject which recognize an epitope of an HCV polypeptide, wherein a population of activated T cells recognizes an epitope of the NS3, NS4, NS5 and/or core polypeptides.

HCV Genome and Recombinant Proteins

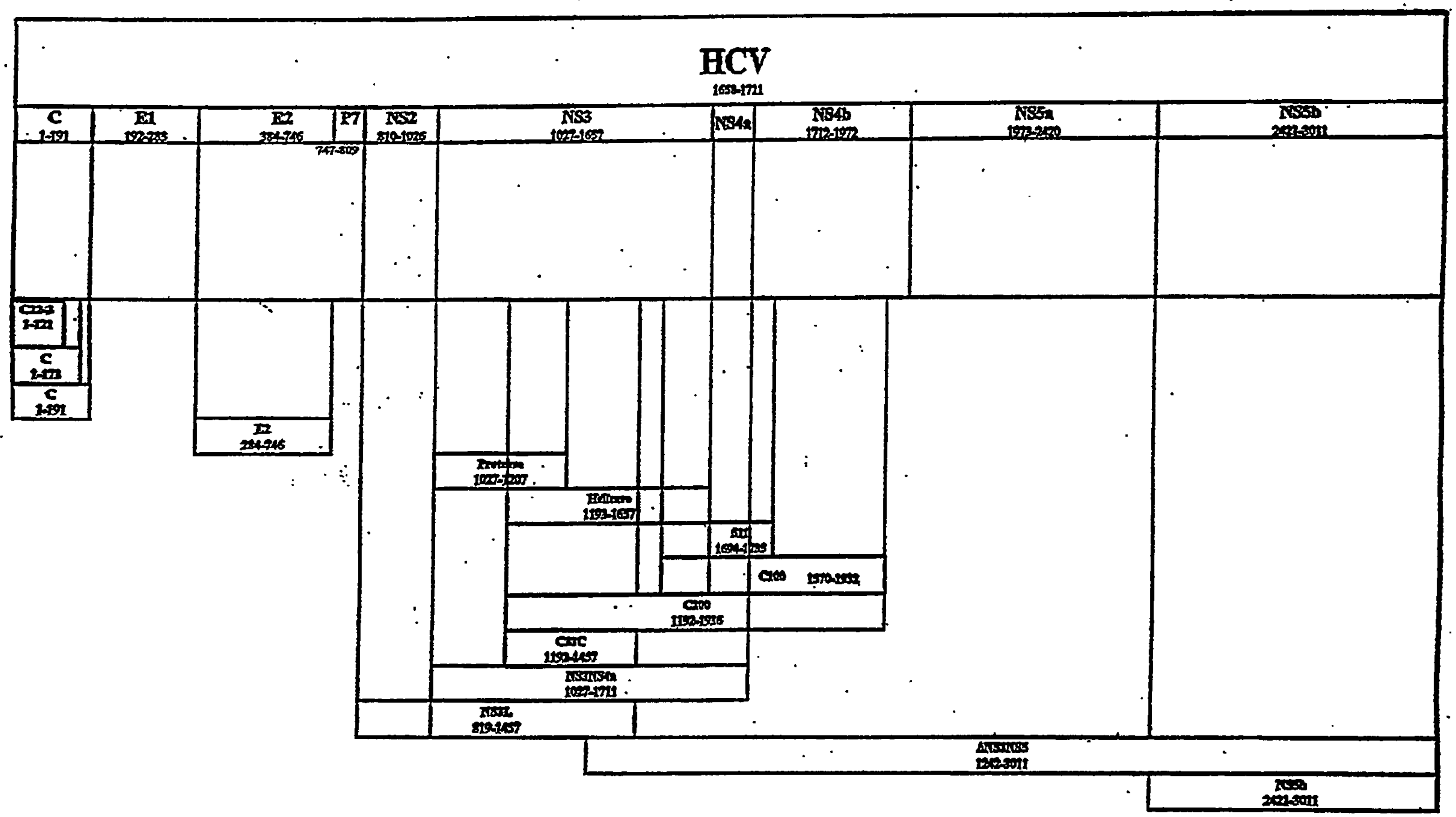


FIG. 1

```
                                1                                  10
                                M   A   P   I   T   A   Y   A   Q   Q
                               ATG GCG CCC ATC ACG GCG TAC GCC CAG CAG

                                                  20
 T   R   G   L   L   G   C   I   I   T   S   L   T   G   R
ACA AGG GGC CTC CTA GGG TGC ATA ATC ACC AGC CTA ACT GGC CGG

                    30                                        40
 D   K   N   Q   V   E   G   E   V   Q   I   V   S   T   A
GAC AAA AAC CAA GTG GAG GGT GAG GTC CAG ATT GTG TCA ACT GCT

                                                  50
 A   Q   T   F   L   A   T   C   I   N   G   V   C   W   T
GCC CAA ACC TTC CTG GCA ACG TGC ATC AAT GGG GTG TGC TGG ACT

                    60                                        70
 V   Y   H   G   A   G   T   R   T   I   A   S   P   K   G
GTC TAC CAC GGG GCC GGA ACG AGG ACC ATC GCG TCA CCC AAG GGT

                                                  80
 P   V   I   Q   M   Y   T   N   V   D   Q   D   L   V   G
CCT GTC ATC CAG ATG TAT ACC AAT GTA GAC CAA GAC CTT GTG GGC

                    90                                       100
 W   P   A   P   Q   G   S   R   S   L   T   P   C   T   C
TGG CCC GCT CCG CAA GGT AGC CGA TCA TTG ACA CCC TGC ACT TGC

                                                 110
 G   S   S   D   L   Y   L   V   T   R   H   A   D   V   I
GGC TCC TCG GAC CTT TAC CTG GTC ACG AGG CAC GCC GAT GTC ATT

                   120                                       130
 P   V   R   R   R   G   D   S   R   G   S   L   L   S   P
CCC GTG CGC CGG CGG GGT GAT AGC AGG GGC AGC CTG CTG TCG CCC

                                                 140
 R   P   I   S   Y   L   K   G   S   S   G   G   P   L   L
CGG CCC ATT TCC TAC TTG AAA GGC TCC TCG GGG GGT CCG CTG TTG

                   150                                       160
 C   P   A   G   H   A   V   G   I   F   R   A   A   V   C
TGC CCC GCG GGG CAC GCC GTG GGC ATA TTT AGG GCC GCG GTG TGC

                                                 170
 T   R   G   V   A   K   A   V   D   F   I   P   V   E   N
ACC CGT GGA GTG GCT AAG GCG GTG GAC TTT ATC CCT GTG GAG AAC

                   180
 L   E   T   T   M   R   S
CTA GAG ACA ACC ATG AGG TCC
```

FIG. 2

MATURE E1

SerPheSerIlePheLeuLeuAlaLeuLeuSerCysLeuThrValProAlaSerAlaTyr 192
TCTTTCTCTATCTTCCTTCTGGCCCTGCTCTCTTGCTTGACTGTGCCCGCTTCGGCCTAC
AGAAAGAGATAGAAGGAAGACCGGGACGAGAGAACGAACTGACACGGGCGAAGCCGGATG

GlnValArgAsnSerThrGlyLeuTyrHisValThrAsnAspCysProAsnSerSerIle 212
CAAGTGCGCAACTCCACGGGGCTCTACCACGTCACCAATGATTGCCCTAACTCGAGTATT
GTTCACGCGTTGAGGTGCCCCGAGATGGTGCAGTGGTTACTAACGGGATTGAGCTCATAA

ValTyrGluAlaAlaAspAlaIleLeuHisThrProGlyCysValProCysValArgGlu 232
GTGTACGAGGCGGCCGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGCGAG
CACATGCTCCGCCGGCTACGGTAGGACGTGTGAGGCCCCACGCAGGGAACGCAAGCGCTC

GlyAsnAlaSerArgCysTrpValAlaMetThrProThrValAlaThrArgAspGlyLys 252
GGCAACGCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGATGGCAAA
CCGTTGCGGAGCTCCACAACCCACCGCTACTGGGGATGCCACCGGTGGTCCCTACCGTTT

LeuProAlaThrGlnLeuArgArgHisIleAspLeuLeuValGlySerAlaThrLeuCys 272
CTCCCCGCGACGCAGCTTCGACGTCACATCGATCTGCTTGTCGGGAGCGCCACCCTCTGT
GAGGGGCGCTGCGTCGAAGCTGCAGTGTAGCTAGACGAACAGCCCTCGCGGTGGGAGACA

SerAlaLeuTyrValGlyAspLeuCysGlySerValPheLeuValGlyGlnLeuPheThr 292
TCGGCCCTCTACGTGGGGGACCTGTGCGGGTCTGTCTTTCTTGTCGGCCAACTGTTTACC
AGCCGGGAGATGCACCCCCTGGACACGCCCAGACAGAAAGAACAGCCGGTTGACAAATGG

PheSerProArgArgHisTrpThrThrGlnGlyCysAsnCysSerIleTyrProGlyHis 312
TTCTCTCCCAGGCGCCACTGGACGACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCAT
AAGAGAGGGTCCGCGGTGACCTGCTGCGTTCCAACGTTAACGAGATAGATAGGGCCGGTA

IleThrGlyHisArgMetAlaTrpAspMetMetMetAsnTrpSerProThrThrAlaLeu 332
ATAACGGGTCACCGCATGGCATGGGATATGATGATGAACTGGTCCCCTACGACGGCGTTG
TATTGCCCAGTGGCGTACCGTACCCTATACTACTACTTGACCAGGGGATGCTGCCGCAAC

ValMetAlaGlnLeuLeuArgIleProGlnAlaIleLeuAspMetIleAlaGlyAlaHis 352
GTAATGGCTCAGCTGCTCCGGATCCCACAAGCCATCTTGGACATGATCGCTGGTGCTCAC
CATTACCGAGTCGACGAGGCCTAGGGTGTTCGGTAGAACCTGTACTAGCGACCACGAGTG

TrpGlyValLeuAlaGlyIleAlaTyrPheSerMetValGlyAsnTrpAlaLysValLeu 372
TGGGGAGTCCTGGCGGGCATAGCGTATTTCTCCATGGTGGGGAACTGGGCGAAGGTCCTG
ACCCCTCAGGACCGCCCGTATCGCATAAAGAGGTACCACCCCTTGACCCGCTTCCAGGAC

E2

ValValLeuLeuLeuPheAlaGlyValAspAlaGluThrHisValThrGlyGlySerAla 392
GTAGTGCTGCTGCTATTTGCCGGCGTCGACGCGGAAACCCACGTCACCGGGGGAAGTGCC
CATCACGACGACGATAAACGGCCGCAGCTGCGCCTTTGGGTGCAGTGGCCCCCTTCACGG

GlyHisThrValSerGlyPheValSerLeuLeuAlaProGlyAlaLysGlnAsnValGln 412
GGCCACACTGTGTCTGGATTTGTTAGCCTCCTCGCACCAGGCGCCAAGCAGAACGTCCAG
CCGGTGTGACACAGACCTAAACAATCGGAGGAGCGTGGTCCGCGGTTCGTCTTGCAGGTC

FIG. 3A

```
LeuIleAsnThrAsnGlySerTrpHisLeuAsnSerThrAlaLeuAsnCysAsnAspSer 432
CTGATCAACACCAACGGCAGTTGGCACCTCAATAGCACGGCCCTGAACTGCAATGATAGC
GACTAGTTGTGGTTGCCGTCAACCGTGGAGTTATCGTGCCGGGACTTGACGTTACTATCG

LeuAsnThrGlyTrpLeuAlaGlyLeuPheTyrHisHisLysPheAsnSerSerGlyCys 452
CTCAACACCGGCTGGTTGGCAGGGCTTTTCTATCACCACAAGTTCAACTCTTCAGGCTGT
GAGTTGTGGCCGACCAACCGTCCCGAAAAGATAGTGGTGTTCAAGTTGAGAAGTCCGACA

ProGluArgLeuAlaSerCysArgProLeuThrAspPheAspGlnGlyTrpGlyProIle 472
CCTGAGAGGCTAGCCAGCTGCCGACCCCTTACCGATTTTGACCAGGGCTGGGGCCCTATC
GGACTCTCCGATCGGTCGACGGCTGGGGAATGGCTAAAACTGGTCCCGACCCCGGGATAG

SerTyrAlaAsnGlySerGlyProAspGlnArgProTyrCysTrpHisTyrProProLys 492
AGTTATGCCAACGGAAGCGGCCCCGACCAGCGCCCCTACTGCTGGCACTACCCCCCAAAA
TCAATACGGTTGCCTTCGCCGGGGCTGGTCGCGGGGATGACGACCGTGATGGGGGGTTTT

ProCysGlyIleValProAlaLysSerValCysGlyProValTyrCysPheThrProSer 512
CCTTGCGGTATTGTGCCCGCGAAGAGTGTGTGTGGTCCGGTATATTGCTTCACTCCCAGC
GGAACGCCATAACACGGGCGCTTCTCACACACACCAGGCCATATAACGAAGTGAGGGTCG

ProValValValGlyThrThrAspArgSerGlyAlaProThrTyrSerTrpGlyGluAsn 532
CCCGTGGTGGTGGGAACGACCGACAGGTCGGGCGCGCCCACCTACAGCTGGGGTGAAAAT
GGGCACCACCACCCTTGCTGGCTGTCCAGCCCGCGCGGGTGGATGTCGACCCCACTTTTA

AspThrAspValPheValLeuAsnAsnThrArgProProLeuGlyAsnTrpPheGlyCys 552
GATACGGACGTCTTCGTCCTTAACAATACCAGGCCACCGCTGGGCAATTGGTTCGGTTGT
CTATGCCTGCAGAAGCAGGAATTGTTATGGTCCGGTGGCGACCCGTTAACCAAGCCAACA

ThrTrpMetAsnSerThrGlyPheThrLysValCysGlyAlaProProCysValIleGly 572
ACCTGGATGAACTCAACTGGATTCACCAAAGTGTGCGGAGCGCCTCCTTGTGTCATCGGA
TGGACCTACTTGAGTTGACCTAAGTGGTTTCACACGCCTCGCGGAGGAACACAGTAGCCT

GlyAlaGlyAsnAsnThrLeuHisCysProThrAspCysPheArgLysHisProAspAla 592
GGGGCGGGCAACAACACCCTGCACTGCCCCACTGATTGCTTCCGCAAGCATCCGGACGCC
CCCCGCCCGTTGTTGTGGGACGTGACGGGGTGACTAACGAAGGCGTTCGTAGGCCTGCGG

ThrTyrSerArgCysGlySerGlyProTrpIleThrProArgCysLeuValAspTyrPro 612
ACATACTCTCGGTGCGGCTCCGGTCCCTGGATCACACCCAGGTGCCTGGTCGACTACCCG
TGTATGAGAGCCACGCCGAGGCCAGGGACCTAGTGTGGGTCCACGGACCAGCTGATGGGC

TyrArgLeuTrpHisTyrProCysThrIleAsnTyrThrIlePheLysIleArgMetTyr 632
TATAGGCTTTGGCATTATCCTTGTACCATCAACTACACTATATTTAAAATCAGGATGTAC
ATATCCGAAACCGTAATAGGAACATGGTAGTTGATGTGATATAAATTTTAGTCCTACATG

ValGlyGlyValGluHisArgLeuGluAlaAlaCysAsnTrpThrArgGlyGluArgCys 652
GTGGGAGGGGTCGAGCACAGGCTGGAAGCTGCCTGCAACTGGACGCGGGGCGAACGTTGC
CACCCTCCCCAGCTCGTGTCCGACCTTCGACGGACGTTGACCTGCGCCCCGCTTGCAACG

AspLeuGluAspArgAspArgSerGluLeuSerProLeuLeuLeuThrThrThrGlnTrp 672
GATCTGGAAGATAGGGACAGGTCCGAGCTCAGCCCGTTACTGCTGACCACTACACAGTGG
CTAGACCTTCTATCCCTGTCCAGGCTCGAGTCGGGCAATGACGACTGGTGATGTGTCACC
```

FIG. 3B

```
GlnValLeuProCysSerPheThrThrLeuProAlaLeuSerThrGlyLeuIleHisLeu 692
CAGGTCCTCCCGTGTTCCTTCACAACCCTGCCAGCCTTGTCCACCGGCCTCATCCACCTC
GTCCAGGAGGGCACAAGGAAGTGTTGGGACGGTCGGAACAGGTGGCCGGAGTAGGTGGAG

HisGlnAsnIleValAspValGlnTyrLeuTyrGlyValGlySerSerIleAlaSerTrp 712
CACCAGAACATTGTGGACGTGCAGTACTTGTACGGGGTGGGGTCAAGCATCGCGTCCTGG
GTGGTCTTGTAACACCTGCACGTCATGAACATGCCCCACCCCAGTTCGTAGCGCAGGACC

AlaIleLysTrpGluTyrValValLeuLeuPheLeuLeuLeuAlaAspAlaArgValCys 732
GCCATTAAGTGGGAGTACGTCGTCCTCCTGTTCCTTCTGCTTGCAGACGCGCGCGTCTGC
CGGTAATTCACCCTCATGCAGCAGGAGGACAAGGAAGACGAACGTCTGCGCGCGCAGACG

                                                    P7
SerCysLeuTrpMetMetLeuLeuIleSerGlnAlaGluAlaAlaLeuGluAsnLeuVal 752
TCCTGCTTGTGGATGATGCTACTCATATCCCAAGCGGAAGCGGCTTTGGAGAACCTCGTA
AGGACGAACACCTACTACGATGAGTATAGGGTTCGCCTTCGCCGAAACCTCTTGGAGCAT

IleLeuAsnAlaAlaSerLeuAlaGlyThrHisGlyLeuValSerPheLeuValPhePhe 772
ATACTTAATGCAGCATCCCTGGCCGGGACGCACGGTCTTGTATCCTTCCTCGTGTTCTTC
TATGAATTACGTCGTAGGGACCGGCCCTGCGTGCCAGAACATAGGAAGGAGCACAAGAAG

CysPheAlaTrpTyrLeuLysGlyLysTrpValProGlyAlaValTyrThrPheTyrGly 792
TGCTTTGCATGGTATCTGAAGGGTAAGTGGGTGCCCGGAGCGGTCTACACCTTCTACGGG
ACGAAACGTACCATAGACTTCCCATTCACCCACGGGCCTCGCCAGATGTGGAAGATGCCC

MetTrpProLeuLeuLeuLeuLeuAlaLeuProGlnArgAlaTyrAlaOC          809
ATGTGGCCTCTCCTCCTGCTCCTGTTGGCGTTGCCCCAGCGGGCGTACGCGTAA
TACACCGGAGAGGAGGACGAGGACAACCGCAACGGGGTCGCCCGCATGCGCATT
```

FIG. 3C

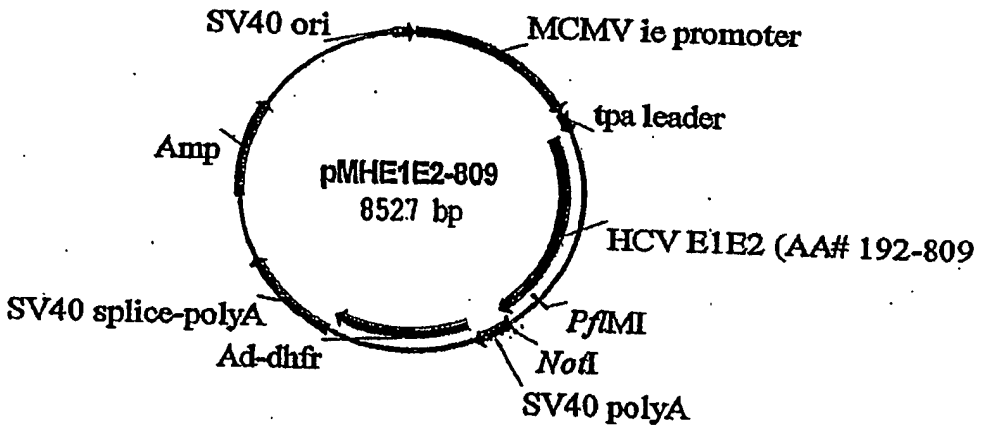

SV40 ori
MCMV ie promoter
tpa leader
Amp
pMHE1E2-809
8527 bp
HCV E1E2 (AA# 192-809
SV40 splice-polyA
PflMI
Ad-dhfr
NotI
SV40 polyA

FIG. 4

```
                  M   A   A   Y   A   A   Q   G   Y   K
                 ATG GCT GCA TAT GCA GCT CAG GGC TAT AAG

                                          20
 V   L   V   L   N   P   S   V   A   A   T   L   G   F   G
GTG CTA GTA CTC AAC CCC TCT GTT GCT GCA ACA CTG GGC TTT GGT

                  30                                      40
 A   Y   M   S   K   A   H   G   I   D   P   N   I   R   T
GCT TAC ATG TCC AAG GCT CAT GGG ATC GAT CCT AAC ATC AGG ACC

                                          50
 G   V   R   T   I   T   T   G   S   P   I   T   Y   S   T
GGG GTG AGA ACA ATT ACC ACT GGC AGC CCC ATC ACG TAC TCC ACC

                  60                                      70
 Y   G   K   F   L   A   D   G   G   C   S   G   G   A   Y
TAC GGC AAG TTC CTT GCC GAC GGC GGG TGC TCG GGG GGC GCT TAT

                                          80
 D   I   I   I   C   D   E   C   H   S   T   D   A   T   S
GAC ATA ATA ATT TGT GAC GAG TGC CAC TCC ACG GAT GCC ACA TCC

                  90                                     100
 I   L   G   I   G   T   V   L   D   Q   A   E   T   A   G
ATC TTG GGC ATT GGC ACT GTC CTT GAC CAA GCA GAG ACT GCG GGG

                                         110
 A   R   L   V   V   L   A   T   A   T   P   P   G   S   V
GCG AGA CTG GTT GTG CTC GCC ACC GCC ACC CCT CCG GGC TCC GTC

                 120                                     130
 T   V   P   H   P   N   I   E   E   V   A   L   S   T   T
ACT GTG CCC CAT CCC AAC ATC GAG GAG GTT GCT CTG TCC ACC ACC

                                         140
 G   E   I   P   F   Y   G   K   A   I   P   L   E   V   I
GGA GAG ATC CCT TTT TAC GGC AAG GCT ATC CCC CTC GAA GTA ATC

                 150                                     160
 K   G   G   R   H   L   I   F   C   H   S   K   K   K   C
AAG GGG GGG AGA CAT CTC ATC TTC TGT CAT TCA AAG AAG AAG TGC

                                         170
 D   E   L   A   A   K   L   V   A   L   G   I   N   A   V
GAC GAA CTC GCC GCA AAG CTG GTC GCA TTG GGC ATC AAT GCC GTG

                 180                                     190
 A   Y   Y   R   G   L   D   V   S   V   I   P   T   S   G
GCC TAC TAC CGC GGT CTT GAC GTG TCC GTC ATC CCG ACC AGC GGC
```

FIG. 5A

```
                                                  200
 D   V   V   V   V   A   T   D   A   L   M   T   G   Y   T
GAT GTT GTC GTC GTG GCA ACC GAT GCC CTC ATG ACC GGC TAT ACC

             210                                      220
 G   D   F   D   S   V   I   D   C   N   T   C   V   T   Q
GGC GAC TTC GAC TCG GTG ATA GAC TGC AAT ACG TGT GTC ACC CAG

                                                  230
 T   V   D   F   S   L   D   P   T   F   T   I   E   T   I
ACA GTC GAT TTC AGC CTT GAC CCT ACC TTC ACC ATT GAG ACA ATC

             240                                      250
 T   L   P   Q   D   A   V   S   R   T   Q   R   R   G   R
ACG CTC CCC CAA GAT GCT GTC TCC CGC ACT CAA CGT CGG GGC AGG

                                                  260
 T   G   R   G   K   P   G   I   Y   R   F   V   A   P   G
ACT GGC AGG GGG AAG CCA GGC ATC TAC AGA TTT GTG GCA CCG GGG

             270                                      280
 E   R   P   S   G   M   F   D   S   S   V   L   C   E   C
GAG CGC CCC TCC GGC ATG TTC GAC TCG TCC GTC CTC TGT GAG TGC

                                                  290
 Y   D   A   G   C   A   W   Y   E   L   T   P   A   E   T
TAT GAC GCA GGC TGT GCT TGG TAT GAG CTC ACG CCC GCC GAG ACT

             300                                      310
 T   V   R   L   R   A   Y   M   N   T   P   G   L   P   V
ACA GTT AGG CTA CGA GCG TAC ATG AAC ACC CCG GGG CTT CCC GTG

                                                  320
 C   Q   D   H   L   E   F   W   E   G   V   F   T   G   L
TGC CAG GAC CAT CTT GAA TTT TGG GAG GGC GTC TTT ACA GGC CTC

             330                                      340
 T   H   I   D   A   H   F   L   S   Q   T   K   Q   S   G
ACT CAT ATA GAT GCC CAC TTT CTA TCC CAG ACA AAG CAG AGT GGG

                                                  350
 E   N   L   P   Y   L   V   A   Y   Q   A   T   V   C   A
GAG AAC CTT CCT TAC CTG GTA GCG TAC CAA GCC ACC GTG TGC GCT

             360                                      370
 R   A   Q   A   P   P   P   S   W   D   Q   M   W   K   C
AGG GCT CAA GCC CCT CCC CCA TCG TGG GAC CAG ATG TGG AAG TGT

                                                  380
 L   I   R   L   K   P   T   L   H   G   P   T   P   L   L
TTG ATT CGC CTC AAG CCC ACC CTC CAT GGG CCA ACA CCC CTG CTA
```

FIG. 5B

```
                  390                                                 400
  Y   R   L   G   A   V   Q   N   E   I   T   L   T   H   P
TAC AGA CTG GGC GCT GTT CAG AAT GAA ATC ACC CTG ACG CAC CCA

                                      410
  V   T   K   Y   I   M   T   C   M   S   A   D   L   E   V
GTC ACC AAA TAC ATC ATG ACA TGC ATG TCG GCC GAC CTG GAG GTC

                  420                                                 430
  V   T   S   T   W   V   L   V   G   G   V   L   A   A   L
GTC ACG AGC ACC TGG GTG CTC GTT GGC GGC GTC CTG GCT GCT TTG

                                      440
  A   A   Y   C   L   S   T   G   C   V   V   I   V   G   R
GCC GCG TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC AGG

                  450                                                 460
  V   V   L   S   G   K   P   A   I   I   P   D   R   E   V
GTC GTC TTG TCC GGG AAG CCG GCA ATC ATA CCT GAC AGG GAA GTC

                                      470
  L   Y   R   E   F   D   E   M   E   E   C   S   Q   H   L
CTC TAC CGA GAG TTC GAT GAG ATG GAA GAG TGC TCT CAG CAC TTA

                  480                                                 490
  P   Y   I   E   Q   G   M   M   L   A   E   Q   F   K   Q
CCG TAC ATC GAG CAA GGG ATG ATG CTC GCC GAG CAG TTC AAG CAG

                                      500
  K   A   L   G   L   L   Q   T   A   S   R   Q   A   E   V
AAG GCC CTC GGC CTC CTG CAG ACC GCG TCC CGT CAG GCA GAG GTT

                  510                                                 520
  I   A   P   A   V   Q   T   N   W   Q   K   L   E   T   F
ATC GCC CCT GCT GTC CAG ACC AAC TGG CAA AAA CTC GAG ACC TTC

                                      530
  W   A   K   H   M   W   N   F   I   S   G   I   Q   Y   L
TGG GCG AAG CAT ATG TGG AAC TTC ATC AGT GGG ATA CAA TAC TTG

                  540                                                 550
  A   G   L   S   T   L   P   G   N   P   A   I   A   S   L
GCG GGC TTG TCA ACG CTG CCT GGT AAC CCC GCC ATT GCT TCA TTG

                                      560
  M   A   F   T   A   A   V   T   S   P   L   T   T   S   Q
ATG GCT TTT ACA GCT GCT GTC ACC AGC CCA CTA ACC ACT AGC CAA

                  570                                                 580
  T   L   L   F   N   I   L   G   G   W   V   A   A   Q   L
ACC CTC CTC TTC AAC ATA TTG GGG GGG TGG GTG GCT GCC CAG CTC
```

FIG. 5C

```
                                                590
 A   A   P   G   A   A   T   A   F   V   G   A   G   L   A
GCC GCC CCC GGT GCC GCT ACT GCC TTT GTG GGC GCT GGC TTA GCT

                600                                     610
 G   A   A   I   G   S   V   G   L   G   K   V   L   I   D
GGC GCC GCC ATC GGC AGT GTT GGA CTG GGG AAG GTC CTC ATA GAC

                                                620
 I   L   A   G   Y   G   A   G   V   A   G   A   L   V   A
ATC CTT GCA GGG TAT GGC GCG GGC GTG GCG GGA GCT CTT GTG GCA

                630                                     640
 F   K   I   M   S   G   E   V   P   S   T   E   D   L   V
TTC AAG ATC ATG AGC GGT GAG GTC CCC TCC ACG GAG GAC CTG GTC

                                                650
 N   L   L   P   A   I   L   S   P   G   A   L   V   V   G
AAT CTA CTG CCC GCC ATC CTC TCG CCC GGA GCC CTC GTA GTC GGC

                660                                     670
 V   V   C   A   A   I   L   R   R   H   V   G   P   G   E
GTG GTC TGT GCA GCA ATA CTG CGC CGG CAC GTT GGC CCG GGC GAG

                                                680
 G   A   V   Q   W   M   N   R   L   I   A   F   A   S   R
GGG GCA GTG CAG TGG ATG AAC CGG CTG ATA GCC TTC GCC TCC CGG

                690                                     700
 G   N   H   V   S   P   T   H   Y   V   P   E   S   D   A
GGG AAC CAT GTT TCC CCC ACG CAC TAC GTG CCG GAG AGC GAT GCA

                                                710
 A   A   R   V   T   A   I   L   S   S   L   T   V   T   Q
GCT GCC CGC GTC ACT GCC ATA CTC AGC AGC CTC ACT GTA ACC CAG

                720                                     730
 L   L   R   R   L   H   Q   W   I   S   S   E   C   T   T
CTC CTG AGG CGA CTG CAC CAG TGG ATA AGC TCG GAG TGT ACC ACT

                                                740
 P   C   S   G   S   W   L   R   D   I   W   D   W   I   C
CCA TGC TCC GGT TCC TGG CTA AGG GAC ATC TGG GAC TGG ATA TGC

                750                                     760
 E   V   L   S   D   F   K   T   W   L   K   A   K   L   M
GAG GTG TTG AGC GAC TTT AAG ACC TGG CTA AAA GCT AAG CTC ATG

                                                770
 P   Q   L   P   G   I   P   F   V   S   C   Q   R   G   Y
CCA CAG CTG CCT GGG ATC CCC TTT GTG TCC TGC CAG CGC GGG TAT
```

FIG. 5D

```
                  780                                                790
 K   G   V   W   R   G   D   G   I   M   H   T   R   C   H
AAG GGG GTC TGG CGA GGG GAC GGC ATC ATG CAC ACT CGC TGC CAC

                                      800
 C   G   A   E   I   T   G   H   V   K   N   G   T   M   R
TGT GGA GCT GAG ATC ACT GGA CAT GTC AAA AAC GGG ACG ATG AGG

                  810                                                820
 I   V   G   P   R   T   C   R   N   M   W   S   G   T   F
ATC GTC GGT CCT AGG ACC TGC AGG AAC ATG TGG AGT GGG ACC TTC

                                      830
 P   I   N   A   Y   T   T   G   P   C   T   P   L   P   A
CCC ATT AAT GCC TAC ACC ACG GGC CCC TGT ACC CCC CTT CCT GCG

                  840                                                850
 P   N   Y   T   F   A   L   W   R   V   S   A   E   E   Y
CCG AAC TAC ACG TTC GCG CTA TGG AGG GTG TCT GCA GAG GAA TAC

                                      860
 V   E   I   R   Q   V   G   D   F   H   Y   V   T   G   M
GTG GAG ATA AGG CAG GTG GGG GAC TTC CAC TAC GTG ACG GGT ATG

                  870                                                880
 T   T   D   N   L   K   C   P   C   Q   V   P   S   P   E
ACT ACT GAC AAT CTT AAA TGC CCG TGC CAG GTC CCA TCG CCC GAA

                                      890
 F   F   T   E   L   D   G   V   R   L   H   R   F   A   P
TTT TTC ACA GAA TTG GAC GGG GTG CGC CTA CAT AGG TTT GCG CCC

                  900                                                910
 P   C   K   P   L   L   R   E   E   V   S   F   R   V   G
CCC TGC AAG CCC TTG CTG CGG GAG GAG GTA TCA TTC AGA GTA GGA

                                      920
 L   H   E   Y   P   V   G   S   Q   L   P   C   E   P   E
CTC CAC GAA TAC CCG GTA GGG TCG CAA TTA CCT TGC GAG CCC GAA

                  930                                                940
 P   D   V   A   V   L   T   S   M   L   T   D   P   S   H
CCG GAC GTG GCC GTG TTG ACG TCC ATG CTC ACT GAT CCC TCC CAT

                                      950
 I   T   A   E   A   A   G   R   R   L   A   R   G   S   P
ATA ACA GCA GAG GCG GCC GGG CGA AGG TTG GCG AGG GGA TCA CCC

                  960                                                970
 P   S   V   A   S   S   S   A   S   Q   L   S   A   P   S
CCC TCT GTG GCC AGC TCC TCG GCT AGC CAG CTA TCC GCT CCA TCT
```

FIG. 5E

```
                                        980
 L   K   A   T   C   T   A   N   H   D   S   P   D   A   E
CTC AAG GCA ACT TGC ACC GCT AAC CAT GAC TCC CCT GAT GCT GAG

                990                                     1000
 L   I   E   A   N   L   L   W   R   Q   E   M   G   G   N
CTC ATA GAG GCC AAC CTC CTA TGG AGG CAG GAG ATG GGC GGC AAC

                                    1010
 I   T   R   V   E   S   E   N   K   V   V   I   L   D   S
ATC ACC AGG GTT GAG TCA GAA AAC AAA GTG GTG ATT CTG GAC TCC

                1020                                    1030
 F   D   P   L   V   A   E   E   D   E   R   E   I   S   V
TTC GAT CCG CTT GTG GCG GAG GAG GAC GAG CGG GAG ATC TCC GTA

                                    1040
 P   A   E   I   L   R   K   S   R   R   F   A   Q   A   L
CCC GCA GAA ATC CTG CGG AAG TCT CGG AGA TTC GCC CAG GCC CTG

                1050                                    1060
 P   V   W   A   R   P   D   Y   N   P   P   L   V   E   T
CCC GTT TGG GCG CGG CCG GAC TAT AAC CCC CCG CTA GTG GAG ACG

                                    1070
 W   K   K   P   D   Y   E   P   P   V   V   H   G   C   P
TGG AAA AAG CCC GAC TAC GAA CCA CCT GTG GTC CAT GGC TGC CCG

                1080                                    1090
 L   P   P   P   K   S   P   P   V   P   P   P   R   K   K
CTT CCA CCT CCA AAG TCC CCT CCT GTG CCT CCG CCT CGG AAG AAG

                                    1100
 R   T   V   V   L   T   E   S   T   L   S   T   A   L   A
CGG ACG GTG GTC CTC ACT GAA TCA ACC CTA TCT ACT GCC TTG GCC

                1110                                    1120
 E   L   A   T   R   S   F   G   S   S   S   T   S   G   I
GAG CTC GCC ACC AGA AGC TTT GGC AGC TCC TCA ACT TCC GGC ATT

                                    1130
 T   G   D   N   T   T   T   S   S   E   P   A   P   S   G
ACG GGC GAC AAT ACG ACA ACA TCC TCT GAG CCC GCC CCT TCT GGC

                1140                                    1150
 C   P   P   D   S   D   A   E   S   Y   S   S   M   P   P
TGC CCC CCC GAC TCC GAC GCT GAG TCC TAT TCC TCC ATG CCC CCC

                                    1160
 L   E   G   E   P   G   D   P   D   L   S   D   G   S   W
CTG GAG GGG GAG CCT GGG GAT CCG GAT CTT AGC GAC GGG TCA TGG
```

FIG. 5F

```
                    1170                                            1180
 S   T   V   S   S   E   A   N   A   E   D   V   V   C   C
TCA ACG GTC AGT AGT GAG GCC AAC GCG GAG GAT GTC GTG TGC TGC

                                        1190
 S   M   S   Y   S   W   T   G   A   L   V   T   P   C   A
TCA ATG TCT TAC TCT TGG ACA GGC GCA CTC GTC ACC CCG TGC GCC

                    1200                                            1210
 A   E   E   Q   K   L   P   I   N   A   L   S   N   S   L
GCG GAA GAA CAG AAA CTG CCC ATC AAT GCA CTA AGC AAC TCG TTG

                                        1220
 L   R   H   H   N   L   V   Y   S   T   T   S   R   S   A
CTA CGT CAC CAC AAT TTG GTG TAT TCC ACC ACC TCA CGC AGT GCT

                    1230                                            1240
 C   Q   R   Q   K   K   V   T   F   D   R   L   Q   V   L
TGC CAA AGG CAG AAG AAA GTC ACA TTT GAC AGA CTG CAA GTT CTG

                                        1250
 D   S   H   Y   Q   D   V   L   K   E   V   K   A   A   A
GAC AGC CAT TAC CAG GAC GTA CTC AAG GAG GTT AAA GCA GCG GCG

                    1260                                            1270
 S   K   V   K   A   N   L   L   S   V   E   E   A   C   S
TCA AAA GTG AAG GCT AAC TTG CTA TCC GTA GAG GAA GCT TGC AGC

                                        1280
 L   T   P   P   H   S   A   K   S   K   F   G   Y   G   A
CTG ACG CCC CCA CAC TCA GCC AAA TCC AAG TTT GGT TAT GGG GCA

                    1290                                            1300
 K   D   V   R   C   H   A   R   K   A   V   T   H   I   N
AAA GAC GTC CGT TGC CAT GCC AGA AAG GCC GTA ACC CAC ATC AAC

                                        1310
 S   V   W   K   D   L   L   E   D   N   V   T   P   I   D
TCC GTG TGG AAA GAC CTT CTG GAA GAC AAT GTA ACA CCA ATA GAC

                    1320                                            1330
 T   T   I   M   A   K   N   E   V   F   C   V   Q   P   E
ACT ACC ATC ATG GCT AAG AAC GAG GTT TTC TGC GTT CAG CCT GAG

                                        1340
 K   G   G   R   K   P   A   R   L   I   V   F   P   D   L
AAG GGG GGT CGT AAG CCA GCT CGT CTC ATC GTG TTC CCC GAT CTG

                    1350                                            1360
 G   V   R   V   C   E   K   M   A   L   Y   D   V   V   T
GGC GTG CGC GTG TGC GAA AAG ATG GCT TTG TAC GAC GTG GTT ACA
```

FIG. 5G

```
                                        1370
 K   L   P   L   A   V   M   G   S   S   Y   G   F   Q   Y
AAG CTC CCC TTG GCC GTG ATG GGA AGC TCC TAC GGA TTC CAA TAC

                1380                                    1390
 S   P   G   Q   R   V   E   F   L   V   Q   A   W   K   S
TCA CCA GGA CAG CGG GTT GAA TTC CTC GTG CAA GCG TGG AAG TCC

                                        1400
 K   K   T   P   M   G   F   S   Y   D   T   R   C   F   D
AAG AAA ACC CCA ATG GGG TTC TCG TAT GAT ACC CGC TGC TTT GAC

                1410                                    1420
 S   T   V   T   E   S   D   I   R   T   E   E   A   I   Y
TCC ACA GTC ACT GAG AGC GAC ATC CGT ACG GAG GAG GCA ATC TAC

                                        1430
 Q   C   C   D   L   D   P   Q   A   R   V   A   I   K   S
CAA TGT TGT GAC CTC GAC CCC CAA GCC CGC GTG GCC ATC AAG TCC

                1440                                    1450
 L   T   E   R   L   Y   V   G   G   P   L   T   N   S   R
CTC ACC GAG AGG CTT TAT GTT GGG GGC CCT CTT ACC AAT TCA AGG

                                        1460
 G   E   N   C   G   Y   R   R   C   R   A   S   G   V   L
GGG GAG AAC TGC GGC TAT CGC AGG TGC CGC GCG AGC GGC GTA CTG

                1470                                    1480
 T   T   S   C   G   N   T   L   T   C   Y   I   K   A   R
ACA ACT AGC TGT GGT AAC ACC CTC ACT TGC TAC ATC AAG GCC CGG

                                        1490
 A   A   C   R   A   A   G   L   Q   D   C   T   M   L   V
GCA GCC TGT CGA GCC GCA GGG CTC CAG GAC TGC ACC ATG CTC GTG

                1500                                    1510
 C   G   D   D   L   V   V   I   C   E   S   A   G   V   Q
TGT GGC GAC GAC TTA GTC GTT ATC TGT GAA AGC GCG GGG GTC CAG

                                        1520
 E   D   A   A   S   L   R   A   F   T   E   A   M   T   R
GAG GAC GCG GCG AGC CTG AGA GCC TTC ACG GAG GCT ATG ACC AGG

                1530                                    1540
 Y   S   A   P   P   G   D   P   P   Q   P   E   Y   D   L
TAC TCC GCC CCC CCT GGG GAC CCC CCA CAA CCA GAA TAC GAC TTG

                                        1550
 E   L   I   T   S   C   S   S   N   V   S   V   A   H   D
GAG CTC ATA ACA TCA TGC TCC TCC AAC GTG TCA GTC GCC CAC GAC
```

FIG. 5H

```
                  1560                                          1570
 G   A   G   K   R   V   Y   Y   L   T   R   D   P   T   T
GGC GCT GGA AAG AGG GTC TAC TAC CTC ACC CGT GAC CCT ACA ACC

                                  1580
 P   L   A   R   A   A   W   E   T   A   R   H   T   P   V
CCC CTC GCG AGA GCT GCG TGG GAG ACA GCA AGA CAC ACT CCA GTC

                  1590                                          1600
 N   S   W   L   G   N   I   I   M   F   A   P   T   L   W
AAT TCC TGG CTA GGC AAC ATA ATC ATG TTT GCC CCC ACA CTG TGG

                                  1610
 A   R   M   I   L   M   T   H   F   F   S   V   L   I   A
GCG AGG ATG ATA CTG ATG ACC CAT TTC TTT AGC GTC CTT ATA GCC

                  1620                                          1630
 R   D   Q   L   E   Q   A   L   D   C   E   I   Y   G   A
AGG GAC CAG CTT GAA CAG GCC CTC GAT TGC GAG ATC TAC GGG GCC

                                  1640
 C   Y   S   I   E   P   L   D   L   P   P   I   I   Q   R
TGC TAC TCC ATA GAA CCA CTG GAT CTA CCT CCA ATC ATT CAA AGA

                  1650                                          1660
 L   H   G   L   S   A   F   S   L   H   S   Y   S   P   G
CTC CAT GGC CTC AGC GCA TTT TCA CTC CAC AGT TAC TCT CCA GGT

                                  1670
 E   I   N   R   V   A   A   C   L   R   K   L   G   V   P
GAA ATC AAT AGG GTG GCC GCA TGC CTC AGA AAA CTT GGG GTA CCG

                  1680                                          1690
 P   L   R   A   W   R   H   R   A   R   S   V   R   A   R
CCC TTG CGA GCT TGG AGA CAC CGG GCC CGG AGC GTC CGC GCT AGG

                                  1700
 L   L   A   R   G   G   R   A   A   I   C   G   K   Y   L
CTT CTG GCC AGA GGA GGC AGG GCT GCC ATA TGT GGC AAG TAC CTC

                  1710                                          1720
 F   N   W   A   V   R   T   K   L   K   L   T   P   I   A
TTC AAC TGG GCA GTA AGA ACA AAG CTC AAA CTC ACT CCA ATA GCG

                                  1730
 A   A   G   Q   L   D   L   S   G   W   F   T   A   G   Y
GCC GCT GGC CAG CTG GAC TTG TCC GGC TGG TTC ACG GCT GGC TAC

                  1740                                          1750
 S   G   G   D   I   Y   H   S   V   S   H   A   R   P   R
AGC GGG GGA GAC ATT TAT CAC AGC GTG TCT CAT GCC CGG CCC CGC
```

FIG. 5I

```
                                        1760
 W   I   W   F   C   L   L   L   L   A   A   G   V   G   I
TGG ATC TGG TTT TGC CTA CTC CTG CTT GCT GCA GGG GTA GGC ATC

                1770                                    1780
 Y   L   L   P   N   R   M   S   T   N   P   K   P   Q   R
TAC CTC CTC CCC AAC CGA ATG AGC ACG AAT CCT AAA CCT CAA AGA

                                        1790
 K   T   K   R   N   T   N   R   R   P   Q   D   V   K   F
AAG ACC AAA CGT AAC ACC AAC CGG CGG CCG CAG GAC GTC AAG TTC

                1800                                    1810
 P   G   G   G   Q   I   V   G   G   V   Y   L   L   P   R
CCG GGT GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG CGC

                                        1820
 R   G   P   R   L   G   V   R   A   T   R   K   T   S   E
AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGA AAG ACT TCC GAG

                1830                                    1840
 R   S   Q   P   R   G   R   R   Q   P   I   P   K   A   R
CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC CCC AAG GCT CGT

                                        1850
 R   P   E   G   R   T   W   A   Q   P   G   Y   P   W   P
CGG CCC GAG GGC AGG ACC TGG GCT CAG CCC GGG TAC CCT TGG CCC

                1860                                    1870
 L   Y   G   N   E   G   C   G   W   A   G   W   L   L   S
CTC TAT GGC AAT GAG GGC TGC GGG TGG GCG GGA TGG CTC CTG TCT

                                        1880
 P   R   G   S   R   P   S   W   G   P   T   D   P   R   R
CCC CGT GGC TCT CGG CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT

                1890    1892
 R   S   R   N   L   G   K
AGG TCG CGC AAT TTG GGT AAG
```

FIG. 5J

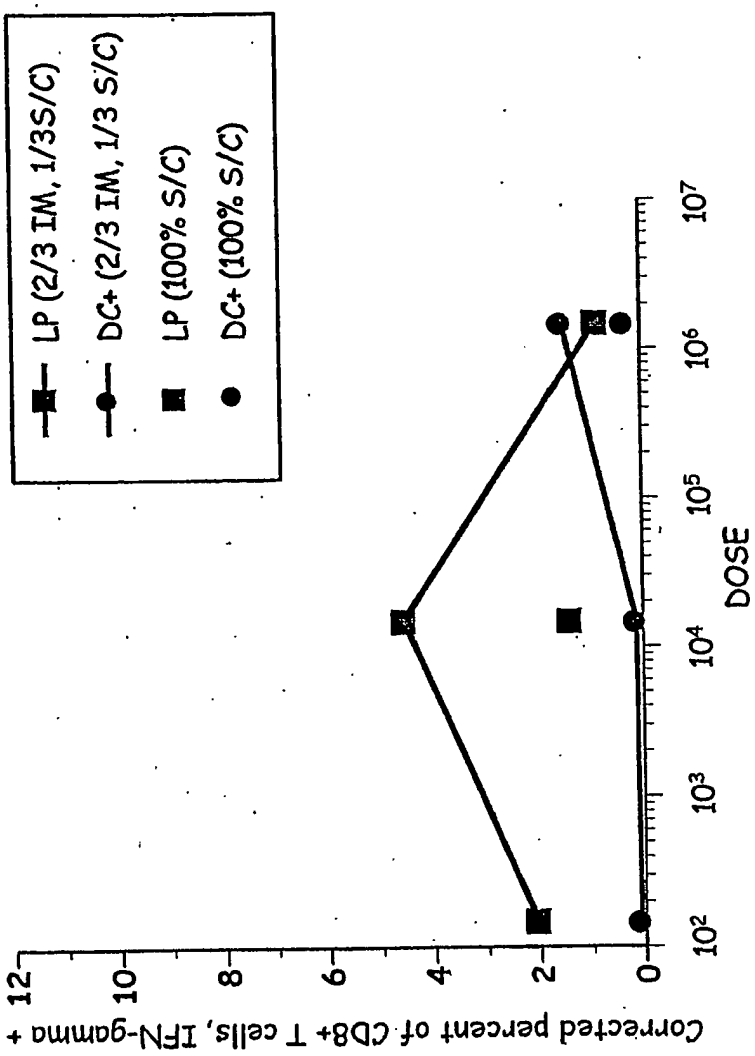

LP (2/3 IM, 1/3S/C)
DC+ (2/3 IM, 1/3 S/C)
LP (100% S/C)
DC+ (100% S/C)
Corrected percent of CD8+ T cells, IFN-gamma +
0
2
4
6
8
10
12
10^2
10^3
10^4
10^5
10^6
10^7
DOSE

FIG. 6

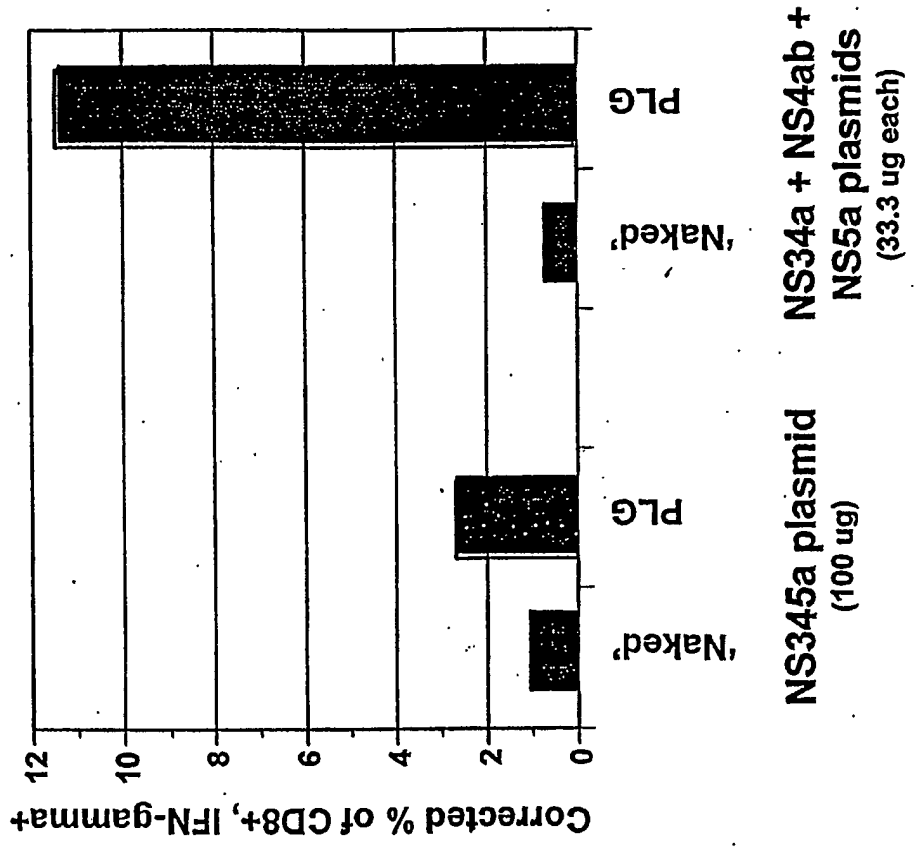

Corrected % of CD8+, IFN-gamma+
0
2
4
6
8
10
12
'Naked'
PLG
'Naked'
PLG
NS345a plasmid
(100 ug)
NS34a + NS4ab +
NS5a plasmids
(33.3 ug each)

FIG. 7

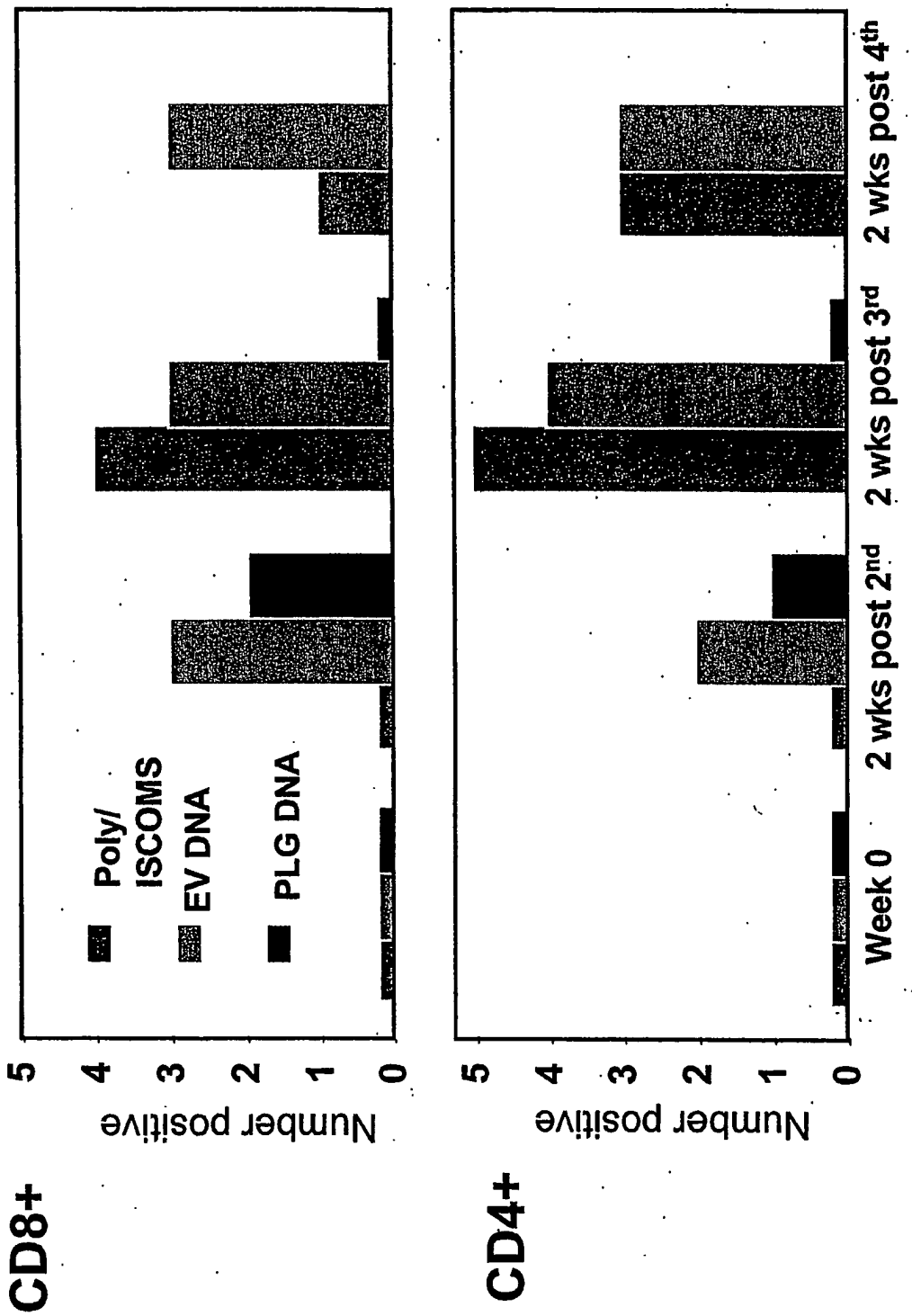

CD8+
Number positive
5
4
3
2
1
0
Poly/ ISCOMS
EV DNA
PLG DNA
CD4+
Number positive
5
4
3
2
1
0
Week 0
2 wks post 2nd
2 wks post 3rd
2 wks post 4th

FIG. 8


CD8+
CD8+ positive
#330
#335
#348
#354
#368
E1
E2
NS3
NS4
NS5a
CD4+
CD4+ positive
#330
#335
#348
#354
#368

FIG. 9

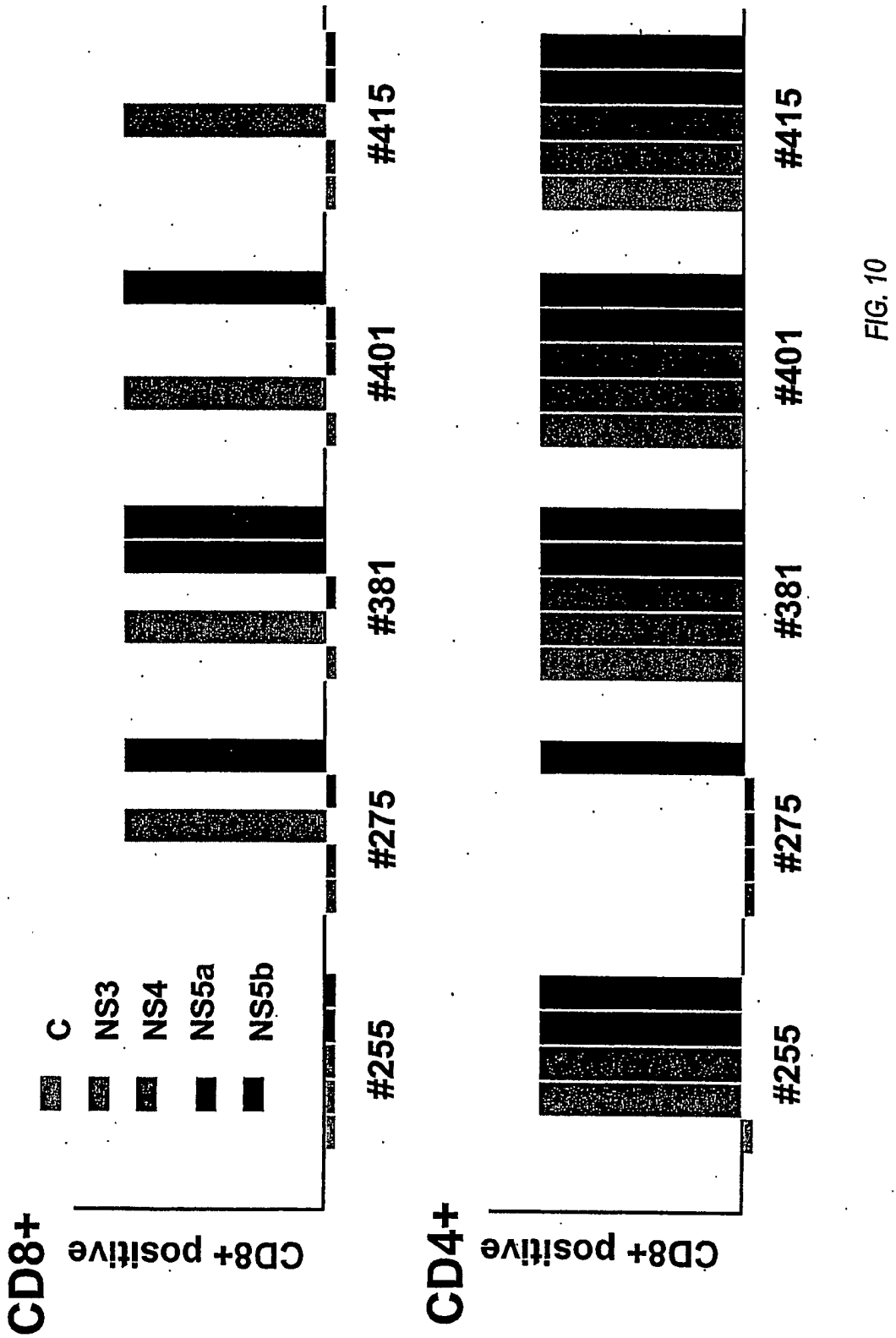

CD8+
CD8+ positive
C
NS3
NS4
NS5a
NS5b
#255
#275
#381
#401
#415
CD4+
CD8+ positive
#255
#275
#381
#401
#415

FIG. 10

SEQUENCE LISTING

<110> HOUGHTON, Michael
COATES, Steve
SELBY, Mark
PALIARD, Xavier

<120> ACTIVATION OF HCV-SPECIFIC T-CELLS

<130> 2300-1612.60

<150> 10/281,341
<151> 2002-10-25

<160> 10

<170> PatentIn Ver. 2.0

<210> 1
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: fusion protein epitope

<400> 1
His Glu Tyr Pro Val Gly Ser Gln Leu
1 5

<210> 2
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: fusion protein epitope

<400> 2
Ala Glu Leu Ile Glu Ala Asn Leu Leu Trp Arg Gln Glu Met Gly
1 5 10 15

<210> 3
<211> 1914
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HCV-1 E1/E2/p7 region

<220>
<221> CDS
<222> (1)..(1911)

<400> 3
tct ttc tct atc ttc ctt ctg gcc ctg ctc tct tgc ttg act gtg ccc 48
Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro

```
  1               5                  10                  15

gct tcg gcc tac caa gtg cgc aac tcc acg ggg ctc tac cac gtc acc   96
Ala Ser Ala Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr
             20                  25                  30

aat gat tgc cct aac tcg agt att gtg tac gag gcg gcc gat gcc atc   144
Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile
         35                  40                  45

ctg cac act ccg ggg tgc gtc cct tgc gtt cgc gag ggc aac gcc tcg   192
Leu His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser
     50                  55                  60

agg tgt tgg gtg gcg atg acc cct acg gtg gcc acc agg gat ggc aaa   240
Arg Cys Trp Val Ala Met Thr Pro Thr Val Ala Thr Arg Asp Gly Lys
 65                  70                  75                  80

ctc ccc gcg acg cag ctt cga cgt cac atc gat ctg ctt gtc ggg agc   288
Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser
                 85                  90                  95

gcc acc ctc tgt tcg gcc ctc tac gtg ggg gac ctg tgc ggg tct gtc   336
Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val
            100                 105                 110

ttt ctt gtc ggc caa ctg ttt acc ttc tct ccc agg cgc cac tgg acg   384
Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr
        115                 120                 125

acg caa ggt tgc aat tgc tct atc tat ccc ggc cat ata acg ggt cac   432
Thr Gln Gly Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His
    130                 135                 140

cgc atg gca tgg gat atg atg atg aac tgg tcc cct acg acg gcg ttg   480
Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu
145                 150                 155                 160

gta atg gct cag ctg ctc cgg atc cca caa gcc atc ttg gac atg atc   528
Val Met Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Leu Asp Met Ile
                165                 170                 175

gct ggt gct cac tgg gga gtc ctg gcg ggc ata gcg tat ttc tcc atg   576
Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met
            180                 185                 190

gtg ggg aac tgg gcg aag gtc ctg gta gtg ctg ctg cta ttt gcc ggc   624
Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly
        195                 200                 205

gtc gac gcg gaa acc cac gtc acc ggg gga agt gcc ggc cac act gtg   672
Val Asp Ala Glu Thr His Val Thr Gly Gly Ser Ala Gly His Thr Val
    210                 215                 220

tct gga ttt gtt agc ctc ctc gca cca ggc gcc aag cag aac gtc cag   720
Ser Gly Phe Val Ser Leu Leu Ala Pro Gly Ala Lys Gln Asn Val Gln
225                 230                 235                 240

ctg atc aac acc aac ggc agt tgg cac ctc aat agc acg gcc ctg aac   768
Leu Ile Asn Thr Asn Gly Ser Trp His Leu Asn Ser Thr Ala Leu Asn
                245                 250                 255
```

```
tgc aat gat agc ctc aac acc ggc tgg ttg gca ggg ctt ttc tat cac   816
Cys Asn Asp Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr His
            260                 265                 270

cac aag ttc aac tct tca ggc tgt cct gag agg cta gcc agc tgc cga   864
His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys Arg
        275                 280                 285

ccc ctt acc gat ttt gac cag ggc tgg ggc cct atc agt tat gcc aac   912
Pro Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn
    290                 295                 300

gga agc ggc ccc gac cag cgc ccc tac tgc tgg cac tac ccc cca aaa   960
Gly Ser Gly Pro Asp Gln Arg Pro Tyr Cys Trp His Tyr Pro Pro Lys
305                 310                 315                 320

cct tgc ggt att gtg ccc gcg aag agt gtg tgt ggt ccg gta tat tgc   1008
Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr Cys
                325                 330                 335

ttc act ccc agc ccc gtg gtg gtg gga acg acc gac agg tcg ggc gcg   1056
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly Ala
            340                 345                 350

ccc acc tac agc tgg ggt gaa aat gat acg gac gtc ttc gtc ctt aac   1104
Pro Thr Tyr Ser Trp Gly Glu Asn Asp Thr Asp Val Phe Val Leu Asn
        355                 360                 365

aat acc agg cca ccg ctg ggc aat tgg ttc ggt tgt acc tgg atg aac   1152
Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met Asn
    370                 375                 380

tca act gga ttc acc aaa gtg tgc gga gcg cct cct tgt gtc atc gga   1200
Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile Gly
385                 390                 395                 400

ggg gcg ggc aac aac acc ctg cac tgc ccc act gat tgc ttc cgc aag   1248
Gly Ala Gly Asn Asn Thr Leu His Cys Pro Thr Asp Cys Phe Arg Lys
                405                 410                 415

cat ccg gac gcc aca tac tct cgg tgc ggc tcc ggt ccc tgg atc aca   1296
His Pro Asp Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile Thr
            420                 425                 430

ccc agg tgc ctg gtc gac tac ccg tat agg ctt tgg cat tat cct tgt   1344
Pro Arg Cys Leu Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys
        435                 440                 445

acc atc aac tac act ata ttt aaa atc agg atg tac gtg gga ggg gtc   1392
Thr Ile Asn Tyr Thr Ile Phe Lys Ile Arg Met Tyr Val Gly Gly Val
    450                 455                 460
gag cac agg ctg gaa gct gcc tgc aac tgg acg cgg ggc gaa cgt tgc   1440
Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys
465                 470                 475                 480

gat ctg gaa gat agg gac agg tcc gag ctc agc ccg tta ctg ctg acc   1488
Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu Thr
                485                 490                 495

act aca cag tgg cag gtc ctc ccg tgt tcc ttc aca acc ctg cca gcc   1536
```

```
Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala
            500                 505                 510

ttg tcc acc ggc ctc atc cac ctc cac cag aac att gtg gac gtg cag   1584
Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val Asp Val Gln
        515                 520                 525

tac ttg tac ggg gtg ggg tca agc atc gcg tcc tgg gcc att aag tgg   1632
Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala Ile Lys Trp
    530                 535                 540

gag tac gtc gtc ctc ctg ttc ctt ctg ctt gca gac gcg cgc gtc tgc   1680
Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val Cys
545                 550                 555                 560

tcc tgc ttg tgg atg atg cta ctc ata tcc caa gcg gaa gcg gct ttg   1728
Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala Ala Leu
                565                 570                 575

gag aac ctc gta ata ctt aat gca gca tcc ctg gcc ggg acg cac ggt   1776
Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly Thr His Gly
            580                 585                 590

ctt gta tcc ttc ctc gtg ttc ttc tgc ttt gca tgg tat ctg aag ggt   1824
Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr Leu Lys Gly
        595                 600                 605

aag tgg gtg ccc gga gcg gtc tac acc ttc tac ggg atg tgg cct ctc   1872
Lys Trp Val Pro Gly Ala Val Tyr Thr Phe Tyr Gly Met Trp Pro Leu
    610                 615                 620

ctc ctg ctc ctg ttg gcg ttg ccc cag cgg gcg tac gcg taa           1914
Leu Leu Leu Leu Leu Ala Leu Pro Gln Arg Ala Tyr Ala
625                 630                 635
```

<210> 4
<211> 637
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: HCV-1 E1/E2/p7 region

<400> 4

```
Ser Phe Ser Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro
  1               5                  10                  15

Ala Ser Ala Tyr Gln Val Arg Asn Ser Thr Gly Leu Tyr His Val Thr
             20                  25                  30

Asn Asp Cys Pro Asn Ser Ser Ile Val Tyr Glu Ala Ala Asp Ala Ile
         35                  40                  45

Leu His Thr Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser
     50                  55                  60

Arg Cys Trp Val Ala Met Thr Pro Thr Val Ala Thr Arg Asp Gly Lys
 65                  70                  75                  80

Leu Pro Ala Thr Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser
```

```
                85                  90                  95

Ala Thr Leu Cys Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val
            100                 105                 110

Phe Leu Val Gly Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr
        115                 120                 125

Thr Gln Gly Cys Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His
    130                 135                 140

Arg Met Ala Trp Asp Met Met Met Asn Trp Ser Pro Thr Thr Ala Leu
145                 150                 155                 160

Val Met Ala Gln Leu Leu Arg Ile Pro Gln Ala Ile Leu Asp Met Ile
                165                 170                 175

Ala Gly Ala His Trp Gly Val Leu Ala Gly Ile Ala Tyr Phe Ser Met
            180                 185                 190

Val Gly Asn Trp Ala Lys Val Leu Val Val Leu Leu Leu Phe Ala Gly
        195                 200                 205

Val Asp Ala Glu Thr His Val Thr Gly Gly Ser Ala Gly His Thr Val
    210                 215                 220

Ser Gly Phe Val Ser Leu Leu Ala Pro Gly Ala Lys Gln Asn Val Gln
225                 230                 235                 240

Leu Ile Asn Thr Asn Gly Ser Trp His Leu Asn Ser Thr Ala Leu Asn
                245                 250                 255

Cys Asn Asp Ser Leu Asn Thr Gly Trp Leu Ala Gly Leu Phe Tyr His
            260                 265                 270

His Lys Phe Asn Ser Ser Gly Cys Pro Glu Arg Leu Ala Ser Cys Arg
        275                 280                 285

Pro Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn
    290                 295                 300

Gly Ser Gly Pro Asp Gln Arg Pro Tyr Cys Trp His Tyr Pro Pro Lys
305                 310                 315                 320

Pro Cys Gly Ile Val Pro Ala Lys Ser Val Cys Gly Pro Val Tyr Cys
                325                 330                 335
Phe Thr Pro Ser Pro Val Val Val Gly Thr Thr Asp Arg Ser Gly Ala
            340                 345                 350

Pro Thr Tyr Ser Trp Gly Glu Asn Asp Thr Asp Val Phe Val Leu Asn
        355                 360                 365

Asn Thr Arg Pro Pro Leu Gly Asn Trp Phe Gly Cys Thr Trp Met Asn
    370                 375                 380

Ser Thr Gly Phe Thr Lys Val Cys Gly Ala Pro Pro Cys Val Ile Gly
385                 390                 395                 400

Gly Ala Gly Asn Asn Thr Leu His Cys Pro Thr Asp Cys Phe Arg Lys
                405                 410                 415
```

```
His Pro Asp Ala Thr Tyr Ser Arg Cys Gly Ser Gly Pro Trp Ile Thr
            420                 425                 430

Pro Arg Cys Leu Val Asp Tyr Pro Tyr Arg Leu Trp His Tyr Pro Cys
        435                 440                 445

Thr Ile Asn Tyr Thr Ile Phe Lys Ile Arg Met Tyr Val Gly Gly Val
    450                 455                 460

Glu His Arg Leu Glu Ala Ala Cys Asn Trp Thr Arg Gly Glu Arg Cys
465                 470                 475                 480

Asp Leu Glu Asp Arg Asp Arg Ser Glu Leu Ser Pro Leu Leu Leu Thr
                485                 490                 495

Thr Thr Gln Trp Gln Val Leu Pro Cys Ser Phe Thr Thr Leu Pro Ala
            500                 505                 510

Leu Ser Thr Gly Leu Ile His Leu His Gln Asn Ile Val Asp Val Gln
        515                 520                 525

Tyr Leu Tyr Gly Val Gly Ser Ser Ile Ala Ser Trp Ala Ile Lys Trp
    530                 535                 540

Glu Tyr Val Val Leu Leu Phe Leu Leu Leu Ala Asp Ala Arg Val Cys
545                 550                 555                 560

Ser Cys Leu Trp Met Met Leu Leu Ile Ser Gln Ala Glu Ala Ala Leu
                565                 570                 575

Glu Asn Leu Val Ile Leu Asn Ala Ala Ser Leu Ala Gly Thr His Gly
            580                 585                 590

Leu Val Ser Phe Leu Val Phe Phe Cys Phe Ala Trp Tyr Leu Lys Gly
        595                 600                 605

Lys Trp Val Pro Gly Ala Val Tyr Thr Phe Tyr Gly Met Trp Pro Leu
    610                 615                 620

Leu Leu Leu Leu Leu Ala Leu Pro Gln Arg Ala Tyr Ala
625                 630                 635


<210> 5
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: consensus sequence

<400> 5
Gly Ser Ala Ala Arg Thr Thr Ser Gly Phe Val Ser Leu Phe Ala Pro
  1               5                  10                  15

Gly Ala Lys Gln Asn
            20


<210> 6
<211> 20
```

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: exemplary CpG oligonucleotide

<400> 6
tccatgacgt tcctgacgtt 20

<210> 7
<211> 5676
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: representative NS345Core fusion protein

<220>
<221> CDS
<222> (1)..(5676)

<400> 7
atg gct gca tat gca gct cag ggc tat aag gtg cta gta ctc aac ccc 48
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
1 5 10 15

tct gtt gct gca aca ctg ggc ttt ggt gct tac atg tcc aag gct cat 96
Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
20 25 30

ggg atc gat cct aac atc agg acc ggg gtg aga aca att acc act ggc 144
Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
35 40 45

agc ccc atc acg tac tcc acc tac ggc aag ttc ctt gcc gac ggc ggg 192
Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
50 55 60
tgc tcg ggg ggc gct tat gac ata ata att tgt gac gag tgc cac tcc 240
Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
65 70 75 80

acg gat gcc aca tcc atc ttg ggc att ggc act gtc ctt gac caa gca 288
Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
85 90 95

gag act gcg ggg gcg aga ctg gtt gtg ctc gcc acc gcc acc cct ccg 336
Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
100 105 110

ggc tcc gtc act gtg ccc cat ccc aac atc gag gag gtt gct ctg tcc 384
Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
115 120 125

acc acc gga gag atc cct ttt tac ggc aag gct atc ccc ctc gaa gta 432
Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
130 135 140

atc aag ggg ggg aga cat ctc atc ttc tgt cat tca aag aag aag tgc 480

```
Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
145                 150                 155                 160

gac gaa ctc gcc gca aag ctg gtc gca ttg ggc atc aat gcc gtg gcc   528
Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
                165                 170                 175

tac tac cgc ggt ctt gac gtg tcc gtc atc ccg acc agc ggc gat gtt   576
Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
            180                 185                 190

gtc gtc gtg gca acc gat gcc ctc atg acc ggc tat acc ggc gac ttc   624
Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
        195                 200                 205

gac tcg gtg ata gac tgc aat acg tgt gtc acc cag aca gtc gat ttc   672
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
    210                 215                 220

agc ctt gac cct acc ttc acc att gag aca atc acg ctc ccc caa gat   720
Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
225                 230                 235                 240

gct gtc tcc cgc act caa cgt cgg ggc agg act ggc agg ggg aag cca   768
Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
                245                 250                 255

ggc atc tac aga ttt gtg gca ccg ggg gag cgc ccc tcc ggc atg ttc   816
Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe
            260                 265                 270

gac tcg tcc gtc ctc tgt gag tgc tat gac gca ggc tgt gct tgg tat   864
Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr
        275                 280                 285

gag ctc acg ccc gcc gag act aca gtt agg cta cga gcg tac atg aac   912
Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn
    290                 295                 300

acc ccg ggg ctt ccc gtg tgc cag gac cat ctt gaa ttt tgg gag ggc   960
Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly
305                 310                 315                 320

gtc ttt aca ggc ctc act cat ata gat gcc cac ttt cta tcc cag aca   1008
Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr
                325                 330                 335

aag cag agt ggg gag aac ctt cct tac ctg gta gcg tac caa gcc acc   1056
Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr
            340                 345                 350

gtg tgc gct agg gct caa gcc cct ccc cca tcg tgg gac cag atg tgg   1104
Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp
        355                 360                 365

aag tgt ttg att cgc ctc aag ccc acc ctc cat ggg cca aca ccc ctg   1152
Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu
    370                 375                 380

cta tac aga ctg ggc gct gtt cag aat gaa atc acc ctg acg cac cca   1200
Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro
```

```
385                 390                 395                 400

gtc acc aaa tac atc atg aca tgc atg tcg gcc gac ctg gag gtc gtc   1248
Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
                405                 410                 415

acg agc acc tgg gtg ctc gtt ggc ggc gtc ctg gct gct ttg gcc gcg   1296
Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
            420                 425                 430

tat tgc ctg tca aca ggc tgc gtg gtc ata gtg ggc agg gtc gtc ttg   1344
Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
        435                 440                 445

tcc ggg aag ccg gca atc ata cct gac agg gaa gtc ctc tac cga gag   1392
Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
    450                 455                 460

ttc gat gag atg gaa gag tgc tct cag cac tta ccg tac atc gag caa   1440
Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
465                 470                 475                 480

ggg atg atg ctc gcc gag cag ttc aag cag aag gcc ctc ggc ctc ctg   1488
Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
                485                 490                 495

cag acc gcg tcc cgt cag gca gag gtt atc gcc cct gct gtc cag acc   1536
Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
            500                 505                 510

aac tgg caa aaa ctc gag acc ttc tgg gcg aag cat atg tgg aac ttc   1584
Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
        515                 520                 525

atc agt ggg ata caa tac ttg gcg ggc ttg tca acg ctg cct ggt aac   1632
Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
    530                 535                 540

ccc gcc att gct tca ttg atg gct ttt aca gct gct gtc acc agc cca   1680
Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
545                 550                 555                 560

cta acc act agc caa acc ctc ctc ttc aac ata ttg ggg ggg tgg gtg   1728
Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
                565                 570                 575

gct gcc cag ctc gcc gcc ccc ggt gcc gct act gcc ttt gtg ggc gct   1776
Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
            580                 585                 590

ggc tta gct ggc gcc gcc atc ggc agt gtt gga ctg ggg aag gtc ctc   1824
Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
        595                 600                 605

ata gac atc ctt gca ggg tat ggc gcg ggc gtg gcg gga gct ctt gtg   1872
Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
    610                 615                 620

gca ttc aag atc atg agc ggt gag gtc ccc tcc acg gag gac ctg gtc   1920
Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
625                 630                 635                 640
```

```
aat cta ctg ccc gcc atc ctc tcg ccc gga gcc ctc gta gtc ggc gtg   1968
Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
                645                 650                 655

gtc tgt gca gca ata ctg cgc cgg cac gtt ggc ccg ggc gag ggg gca   2016
Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
            660                 665                 670

gtg cag tgg atg aac cgg ctg ata gcc ttc gcc tcc cgg ggg aac cat   2064
Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
        675                 680                 685

gtt tcc ccc acg cac tac gtg ccg gag agc gat gca gct gcc cgc gtc   2112
Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
    690                 695                 700

act gcc ata ctc agc agc ctc act gta acc cag ctc ctg agg cga ctg   2160
Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
705                 710                 715                 720

cac cag tgg ata agc tcg gag tgt acc act cca tgc tcc ggt tcc tgg   2208
His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
                725                 730                 735

cta agg gac atc tgg gac tgg ata tgc gag gtg ttg agc gac ttt aag   2256
Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
            740                 745                 750

acc tgg cta aaa gct aag ctc atg cca cag ctg cct ggg atc ccc ttt   2304
Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
        755                 760                 765

gtg tcc tgc cag cgc ggg tat aag ggg gtc tgg cga ggg gac ggc atc   2352
Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
    770                 775                 780

atg cac act cgc tgc cac tgt gga gct gag atc act gga cat gtc aaa   2400
Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
785                 790                 795                 800

aac ggg acg atg agg atc gtc ggt cct agg acc tgc agg aac atg tgg   2448
Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
                805                 810                 815

agt ggg acc ttc ccc att aat gcc tac acc acg ggc ccc tgt acc ccc   2496
Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
            820                 825                 830

ctt cct gcg ccg aac tac acg ttc gcg cta tgg agg gtg tct gca gag   2544
Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
        835                 840                 845

gaa tac gtg gag ata agg cag gtg ggg gac ttc cac tac gtg acg ggt   2592
Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
    850                 855                 860

atg act act gac aat ctt aaa tgc ccg tgc cag gtc cca tcg ccc gaa   2640
Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
865                 870                 875                 880
```

```
ttt ttc aca gaa ttg gac ggg gtg cgc cta cat agg ttt gcg ccc ccc   2688
Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
                885                 890                 895

tgc aag ccc ttg ctg cgg gag gag gta tca ttc aga gta gga ctc cac   2736
Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
            900                 905                 910

gaa tac ccg gta ggg tcg caa tta cct tgc gag ccc gaa ccg gac gtg   2784
Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
        915                 920                 925

gcc gtg ttg acg tcc atg ctc act gat ccc tcc cat ata aca gca gag   2832
Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
    930                 935                 940

gcg gcc ggg cga agg ttg gcg agg gga tca ccc ccc tct gtg gcc agc   2880
Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
945                 950                 955                 960

tcc tcg gct agc cag cta tcc gct cca tct ctc aag gca act tgc acc   2928
Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
                965                 970                 975

gct aac cat gac tcc cct gat gct gag ctc ata gag gcc aac ctc cta   2976
Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
            980                 985                 990

tgg agg cag gag atg ggc ggc aac atc acc agg gtt gag tca gaa aac   3024
Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
        995                1000                1005

aaa gtg gtg att ctg gac tcc ttc gat ccg ctt gtg gcg gag gag gac   3072
Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
   1010                1015                1020

gag cgg gag atc tcc gta ccc gca gaa atc ctg cgg aag tct cgg aga   3120
Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
1025                1030                1035                1040

ttc gcc cag gcc ctg ccc gtt tgg gcg cgg ccg gac tat aac ccc ccg   3168
Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
                1045                1050                1055

cta gtg gag acg tgg aaa aag ccc gac tac gaa cca cct gtg gtc cat   3216
Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
            1060                1065                1070

ggc tgc ccg ctt cca cct cca aag tcc cct cct gtg cct ccg cct cgg   3264
Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
        1075                1080                1085

aag aag cgg acg gtg gtc ctc act gaa tca acc cta tct act gcc ttg   3312
Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
   1090                1095                1100

gcc gag ctc gcc acc aga agc ttt ggc agc tcc tca act tcc ggc att   3360
Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
1105                1110                1115                1120

acg ggc gac aat acg aca aca tcc tct gag ccc gcc cct tct ggc tgc   3408
```

```
Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
                1125                1130                1135

ccc ccc gac tcc gac gct gag tcc tat tcc tcc atg ccc ccc ctg gag   3456
Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
            1140                1145                1150

ggg gag cct ggg gat ccg gat ctt agc gac ggg tca tgg tca acg gtc   3504
Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
        1155                1160                1165

agt agt gag gcc aac gcg gag gat gtc gtg tgc tgc tca atg tct tac   3552
Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
    1170                1175                1180

tct tgg aca ggc gca ctc gtc acc ccg tgc gcc gcg gaa gaa cag aaa   3600
Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
1185                1190                1195                1200

ctg ccc atc aat gca cta agc aac tcg ttg cta cgt cac cac aat ttg   3648
Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
                1205                1210                1215

gtg tat tcc acc acc tca cgc agt gct tgc caa agg cag aag aaa gtc   3696
Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
            1220                1225                1230

aca ttt gac aga ctg caa gtt ctg gac agc cat tac cag gac gta ctc   3744
Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
        1235                1240                1245

aag gag gtt aaa gca gcg gcg tca aaa gtg aag gct aac ttg cta tcc   3792
Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
    1250                1255                1260

gta gag gaa gct tgc agc ctg acg ccc cca cac tca gcc aaa tcc aag   3840
Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
1265                1270                1275                1280

ttt ggt tat ggg gca aaa gac gtc cgt tgc cat gcc aga aag gcc gta   3888
Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
                1285                1290                1295

acc cac atc aac tcc gtg tgg aaa gac ctt ctg gaa gac aat gta aca   3936
Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
            1300                1305                1310

cca ata gac act acc atc atg gct aag aac gag gtt ttc tgc gtt cag   3984
Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
        1315                1320                1325

cct gag aag ggg ggt cgt aag cca gct cgt ctc atc gtg ttc ccc gat   4032
Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
    1330                1335                1340

ctg ggc gtg cgc gtg tgc gaa aag atg gct ttg tac gac gtg gtt aca   4080
Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
1345                1350                1355                1360

aag ctc ccc ttg gcc gtg atg gga agc tcc tac gga ttc caa tac tca   4128
Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
```

```
                1365                1370                1375

cca gga cag cgg gtt gaa ttc ctc gtg caa gcg tgg aag tcc aag aaa   4176
Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
          1380                1385                1390

acc cca atg ggg ttc tcg tat gat acc cgc tgc ttt gac tcc aca gtc   4224
Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
      1395                1400                1405

act gag agc gac atc cgt acg gag gag gca atc tac caa tgt tgt gac   4272
Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
   1410                1415                1420

ctc gac ccc caa gcc cgc gtg gcc atc aag tcc ctc acc gag agg ctt   4320
Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
1425                1430                1435                1440

tat gtt ggg ggc cct ctt acc aat tca agg ggg gag aac tgc ggc tat   4368
Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
                1445                1450                1455

cgc agg tgc cgc gcg agc ggc gta ctg aca act agc tgt ggt aac acc   4416
Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
          1460                1465                1470

ctc act tgc tac atc aag gcc cgg gca gcc tgt cga gcc gca ggg ctc   4464
Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
      1475                1480                1485

cag gac tgc acc atg ctc gtg tgt ggc gac gac tta gtc gtt atc tgt   4512
Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
   1490                1495                1500

gaa agc gcg ggg gtc cag gag gac gcg gcg agc ctg aga gcc ttc acg   4560
Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr
1505                1510                1515                1520

gag gct atg acc agg tac tcc gcc ccc cct ggg gac ccc cca caa cca   4608
Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
                1525                1530                1535

gaa tac gac ttg gag ctc ata aca tca tgc tcc tcc aac gtg tca gtc   4656
Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
          1540                1545                1550

gcc cac gac ggc gct gga aag agg gtc tac tac ctc acc cgt gac cct   4704
Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
      1555                1560                1565

aca acc ccc ctc gcg aga gct gcg tgg gag aca gca aga cac act cca   4752
Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
   1570                1575                1580

gtc aat tcc tgg cta ggc aac ata atc atg ttt gcc ccc aca ctg tgg   4800
Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
1585                1590                1595                1600

gcg agg atg ata ctg atg acc cat ttc ttt agc gtc ctt ata gcc agg   4848
Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
                1605                1610                1615
```

```
gac cag ctt gaa cag gcc ctc gat tgc gag atc tac ggg gcc tgc tac   4896
Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
          1620                1625                1630

tcc ata gaa cca ctg gat cta cct cca atc att caa aga ctc cat ggc   4944
Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
      1635                1640                1645

ctc agc gca ttt tca ctc cac agt tac tct cca ggt gaa atc aat agg   4992
Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
   1650                1655                1660

gtg gcc gca tgc ctc aga aaa ctt ggg gta ccg ccc ttg cga gct tgg   5040
Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
1665                1670                1675                1680

aga cac cgg gcc cgg agc gtc cgc gct agg ctt ctg gcc aga gga ggc   5088
Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
               1685                1690                1695

agg gct gcc ata tgt ggc aag tac ctc ttc aac tgg gca gta aga aca   5136
Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
          1700                1705                1710

aag ctc aaa ctc act cca ata gcg gcc gct ggc cag ctg gac ttg tcc   5184
Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser
      1715                1720                1725

ggc tgg ttc acg gct ggc tac agc ggg gga gac att tat cac agc gtg   5232
Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
   1730                1735                1740

tct cat gcc cgg ccc cgc tgg atc tgg ttt tgc cta ctc ctg ctt gct   5280
Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala
1745                1750                1755                1760

gca ggg gta ggc atc tac ctc ctc ccc aac cga atg agc acg aat cct   5328
Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
               1765                1770                1775

aaa cct caa aga aag acc aaa cgt aac acc aac cgg cgg ccg cag gac   5376
Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
          1780                1785                1790

gtc aag ttc ccg ggt ggc ggt cag atc gtt ggt gga gtt tac ttg ttg   5424
Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
      1795                1800                1805

ccg cgc agg ggc cct aga ttg ggt gtg cgc gcg acg aga aag act tcc   5472
Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
   1810                1815                1820

gag cgg tcg caa cct cga ggt aga cgt cag cct atc ccc aag gct cgt   5520
Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
1825                1830                1835                1840

cgg ccc gag ggc agg acc tgg gct cag ccc ggg tac cct tgg ccc ctc   5568
Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
               1845                1850                1855
```

```
tat ggc aat gag ggc tgc ggg tgg gcg gga tgg ctc ctg tct ccc cgt   5616
Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
          1860                1865                1870

ggc tct cgg cct agc tgg ggc ccc aca gac ccc cgg cgt agg tcg cgc   5664
Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
       1875                1880                1885

aat ttg ggt aag                                                  5676
Asn Leu Gly Lys
   1890
```

<210> 8
<211> 1892
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: representative NS345Core fusion protein

<400> 8

```
Met Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro
  1               5                  10                  15

Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His
            20                  25                  30

Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly
         35                  40                  45

Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly
     50                  55                  60

Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser
 65                  70                  75                  80

Thr Asp Ala Thr Ser Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala
                 85                  90                  95

Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro
            100                 105                 110

Gly Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser
        115                 120                 125

Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Leu Glu Val
    130                 135                 140

Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys
145                 150                 155                 160

Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile Asn Ala Val Ala
                165                 170                 175

Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val
            180                 185                 190

Val Val Val Ala Thr Asp Ala Leu Met Thr Gly Tyr Thr Gly Asp Phe
        195                 200                 205
```

```
Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe
    210                 215                 220

Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Ile Thr Leu Pro Gln Asp
225                 230                 235                 240

Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly Arg Gly Lys Pro
                245                 250                 255

Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro Ser Gly Met Phe
            260                 265                 270

Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr
        275                 280                 285

Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg Ala Tyr Met Asn
    290                 295                 300

Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Gly
305                 310                 315                 320

Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr
                325                 330                 335

Lys Gln Ser Gly Glu Asn Leu Pro Tyr Leu Val Ala Tyr Gln Ala Thr
            340                 345                 350

Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp
        355                 360                 365

Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu
    370                 375                 380

Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Ile Thr Leu Thr His Pro
385                 390                 395                 400

Val Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp Leu Glu Val Val
                405                 410                 415

Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala
            420                 425                 430

Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly Arg Val Val Leu
        435                 440                 445

Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu
    450                 455                 460

Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro Tyr Ile Glu Gln
465                 470                 475                 480

Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala Leu Gly Leu Leu
                485                 490                 495

Gln Thr Ala Ser Arg Gln Ala Glu Val Ile Ala Pro Ala Val Gln Thr
            500                 505                 510

Asn Trp Gln Lys Leu Glu Thr Phe Trp Ala Lys His Met Trp Asn Phe
        515                 520                 525
```

```
Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn
    530                 535                 540

Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala Val Thr Ser Pro
545                 550                 555                 560

Leu Thr Thr Ser Gln Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val
                565                 570                 575

Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala Phe Val Gly Ala
            580                 585                 590

Gly Leu Ala Gly Ala Ala Ile Gly Ser Val Gly Leu Gly Lys Val Leu
        595                 600                 605

Ile Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala Gly Ala Leu Val
    610                 615                 620

Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr Glu Asp Leu Val
625                 630                 635                 640

Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu Val Val Gly Val
                645                 650                 655

Val Cys Ala Ala Ile Leu Arg Arg His Val Gly Pro Gly Glu Gly Ala
            660                 665                 670

Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser Arg Gly Asn His
        675                 680                 685

Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala Ala Ala Arg Val
    690                 695                 700

Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu Leu Arg Arg Leu
705                 710                 715                 720

His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys Ser Gly Ser Trp
                725                 730                 735

Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu Ser Asp Phe Lys
            740                 745                 750

Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro Gly Ile Pro Phe
        755                 760                 765

Val Ser Cys Gln Arg Gly Tyr Lys Gly Val Trp Arg Gly Asp Gly Ile
    770                 775                 780

Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr Gly His Val Lys
785                 790                 795                 800

Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys Arg Asn Met Trp
                805                 810                 815

Ser Gly Thr Phe Pro Ile Asn Ala Tyr Thr Thr Gly Pro Cys Thr Pro
            820                 825                 830

Leu Pro Ala Pro Asn Tyr Thr Phe Ala Leu Trp Arg Val Ser Ala Glu
        835                 840                 845

Glu Tyr Val Glu Ile Arg Gln Val Gly Asp Phe His Tyr Val Thr Gly
```

```
    850                 855                 860

Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Val Pro Ser Pro Glu
865                 870                 875                 880

Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg Phe Ala Pro Pro
                885                 890                 895
Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg Val Gly Leu His
            900                 905                 910

Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro Glu Pro Asp Val
        915                 920                 925

Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His Ile Thr Ala Glu
    930                 935                 940

Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro Ser Val Ala Ser
945                 950                 955                 960

Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys Ala Thr Cys Thr
                965                 970                 975

Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu Ala Asn Leu Leu
            980                 985                 990

Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val Glu Ser Glu Asn
        995                1000                1005

Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val Ala Glu Glu Asp
   1010                1015                1020

Glu Arg Glu Ile Ser Val Pro Ala Glu Ile Leu Arg Lys Ser Arg Arg
1025                1030                1035                1040

Phe Ala Gln Ala Leu Pro Val Trp Ala Arg Pro Asp Tyr Asn Pro Pro
                1045                1050                1055

Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro Pro Val Val His
            1060                1065                1070

Gly Cys Pro Leu Pro Pro Pro Lys Ser Pro Pro Val Pro Pro Pro Arg
        1075                1080                1085

Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu Ser Thr Ala Leu
   1090                1095                1100

Ala Glu Leu Ala Thr Arg Ser Phe Gly Ser Ser Ser Thr Ser Gly Ile
1105                1110                1115                1120

Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys
                1125                1130                1135

Pro Pro Asp Ser Asp Ala Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu
            1140                1145                1150

Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val
        1155                1160                1165

Ser Ser Glu Ala Asn Ala Glu Asp Val Val Cys Cys Ser Met Ser Tyr
   1170                1175                1180
```

Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala Glu Glu Gln Lys
1185 1190 1195 1200

Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu
1205 1210 1215

Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg Gln Lys Lys Val
1220 1225 1230

Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr Gln Asp Val Leu
1235 1240 1245

Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala Asn Leu Leu Ser
1250 1255 1260

Val Glu Glu Ala Cys Ser Leu Thr Pro Pro His Ser Ala Lys Ser Lys
1265 1270 1275 1280

Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala Arg Lys Ala Val
1285 1290 1295

Thr His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu Asp Asn Val Thr
1300 1305 1310

Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
1315 1320 1325

Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
1330 1335 1340

Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Thr
1345 1350 1355 1360

Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser
1365 1370 1375

Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp Lys Ser Lys Lys
1380 1385 1390

Thr Pro Met Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val
1395 1400 1405

Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp
1410 1415 1420

Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu
1425 1430 1435 1440

Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr
1445 1450 1455

Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
1460 1465 1470

Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
1475 1480 1485

Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
1490 1495 1500

Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe Thr

1505 1510 1515 1520
Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Gln Pro
1525 1530 1535

Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val
1540 1545 1550

Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro
1555 1560 1565

Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro
1570 1575 1580

Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala Pro Thr Leu Trp
1585 1590 1595 1600

Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val Leu Ile Ala Arg
1605 1610 1615

Asp Gln Leu Glu Gln Ala Leu Asp Cys Glu Ile Tyr Gly Ala Cys Tyr
1620 1625 1630

Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln Arg Leu His Gly
1635 1640 1645

Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg
1650 1655 1660

Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Ala Trp
1665 1670 1675 1680

Arg His Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ala Arg Gly Gly
1685 1690 1695

Arg Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr
1700 1705 1710

Lys Leu Lys Leu Thr Pro Ile Ala Ala Ala Gly Gln Leu Asp Leu Ser
1715 1720 1725

Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Val
1730 1735 1740

Ser His Ala Arg Pro Arg Trp Ile Trp Phe Cys Leu Leu Leu Leu Ala
1745 1750 1755 1760

Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg Met Ser Thr Asn Pro
1765 1770 1775

Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp
1780 1785 1790

Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu
1795 1800 1805

Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser
1810 1815 1820

Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg
1825 1830 1835 1840

```
Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro Leu
               1845                1850                1855

Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro Arg
           1860                1865                1870

Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser Arg
       1875                1880                1885

Asn Leu Gly Lys
   1890
```

<210> 9
<211> 546
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: representative native NS3 protease domain

<220>
<221> CDS
<222> (1)..(546)

<400> 9

```
atg gcg ccc atc acg gcg tac gcc cag cag aca agg ggc ctc cta ggg   48
Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly
  1               5                  10                  15

tgc ata atc acc agc cta act ggc cgg gac aaa aac caa gtg gag ggt   96
Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
             20                  25                  30

gag gtc cag att gtg tca act gct gcc caa acc ttc ctg gca acg tgc   144
Glu Val Gln Ile Val Ser Thr Ala Ala Gln Thr Phe Leu Ala Thr Cys
         35                  40                  45

atc aat ggg gtg tgc tgg act gtc tac cac ggg gcc gga acg agg acc   192
Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr
     50                  55                  60

atc gcg tca ccc aag ggt cct gtc atc cag atg tat acc aat gta gac   240
Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp
 65                  70                  75                  80

caa gac ctt gtg ggc tgg ccc gct ccg caa ggt agc cga tca ttg aca   288
Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr
                 85                  90                  95

ccc tgc act tgc ggc tcc tcg gac ctt tac ctg gtc acg agg cac gcc   336
Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
            100                 105                 110

gat gtc att ccc gtg cgc cgg cgg ggt gat agc agg ggc agc ctg ctg   384
Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
        115                 120                 125

tcg ccc cgg ccc att tcc tac ttg aaa ggc tcc tcg ggg ggt ccg ctg   432
Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
```

```
    130                 135                 140

ttg tgc ccc gcg ggg cac gcc gtg ggc ata ttt agg gcc gcg gtg tgc   480
Leu Cys Pro Ala Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
145                 150                 155                 160

acc cgt gga gtg gct aag gcg gtg gac ttt atc cct gtg gag aac cta   528
Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu
                165                 170                 175

gag aca acc atg agg tcc                                          546
Glu Thr Thr Met Arg Ser
            180
```

<210> 10
<211> 182
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: representative native NS3 protease domain

<400> 10

```
Met Ala Pro Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly
  1               5                  10                  15

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly
             20                  25                  30

Glu Val Gln Ile Val Ser Thr Ala Ala Gln Thr Phe Leu Ala Thr Cys
         35                  40                  45

Ile Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr
     50                  55                  60

Ile Ala Ser Pro Lys Gly Pro Val Ile Gln Met Tyr Thr Asn Val Asp
 65                  70                  75                  80

Gln Asp Leu Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr
                 85                  90                  95

Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
            100                 105                 110

Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
        115                 120                 125

Ser Pro Arg Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
    130                 135                 140

Leu Cys Pro Ala Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys
145                 150                 155                 160

Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu
                165                 170                 175

Glu Thr Thr Met Arg Ser
            180
```